Vol. 39, Parts 1 and 2 May 1952

BIOMETRIKA

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR
E. S. PEARSON

ASSOCIATE EDITORS
M. G. KENDALL JOHN WISHART

in consultation with
HARALD CRAMÉR R. C. GEARY
J. B. S. HALDANE

Reprinted by offset-litho 1963

ISSUED BY
THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON
PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

[*Issued 4 June* 1952]

BIOMETRIKA PUBLICATIONS: BOOKS OF TABLES

Issued by the Cambridge University Press, Bentley House, London, N.W. 1
and obtainable from any bookseller

Tables of Complete and Incomplete B-Function

EDITED BY KARL PEARSON

59 pages of Introduction and 494 pages of Tables

Price: 55s. net

Tables of the Incomplete Γ-Function

EDITED BY KARL PEARSON

31 pages of Introduction and 164 pages of Tables

Price: 42s. net

Tables of the Complete and Incomplete Elliptic Integrals

(from LEGENDRE'S *Traité des Fonctions Elliptiques*. With autographed portrait of LEGENDRE)

39 pages of Introduction by KARL PEARSON and 94 pages of Tables

Price: 12s. 6d. net

Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples

BY F. N. DAVID

38 pages of Introduction, 55 pages of Tables, 10 Diagrams and 4 Charts

Price: 10s. net

NEW PUBLICATION

Biometrika Tables for Statisticians

The two volumes of *Tables for Statisticians and Biometricians* are now out of print and will not be re-issued.

At the request of the Biometrika Trustees a complete recasting of these Tables has been undertaken by Professor E. S. PEARSON and Dr H. O. HARTLEY. Many of the old tables will be set aside or modified, tables which have been published during the last fifteen years in *Biometrika* will be reproduced and some new tables will be added.

Volume I of the new series, which will include the statistical and auxiliary mathematical tables in more common use is now ready for Press. It will contain 54 tables covering about 135 pp. and an Introduction.

NEW STATISTICAL TABLES: SEPARATES RE-ISSUED FROM BIOMETRIKA

To be obtained from

THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON, W.C.1

I. From *Biometrika*, Vols. 22 and 27

Tests of Normality. By E. S. PEARSON and R. C. GEARY

Price One Shilling, post free

II. From *Biometrika*, Vol. 32, Part 2, pp. 168–181 and 188–189

(1) Table of percentage points of the incomplete beta-function

(2) Table of percentage points of the χ^2 distribution

Stitched together with introductory matter. Price Two Shillings and Sixpence, post free

III. From *Biometrika*, Vol. 32, Parts 3 and 4, pp. 300–310

(1) Table of the probability integral of the range in samples from a normal population

(2) Table of the percentage points of the range

(3) Table of the percentage points of the t-distribution

Stitched together with introductory matter. Price Two Shillings and Sixpence, post free

IV. From *Biometrika*, Vol. 33, Part 1, pp. 73–88

Table of percentage points of the inverted beta (F) distribution

With introductory matter. Price Two Shillings and Sixpence, post free

V. From *Biometrika*, Vol. 33, Part 3, pp. 252–265

(1) Table of the probability integral of the mean deviation in samples from a normal population

(2) Table of the percentage points of the mean deviation

Stitched together with introductory matter. Price Two Shillings and Sixpence, post free

VI. From *Biometrika*, Vol. 33, Part 4, pp. 296–304

Table for testing the homogeneity of a set of estimated variances

With introductory matter. Price Two Shillings, post free

VII. From *Biometrika*, Vol. 35, Parts 1 and 2, pp. 145–156

Table of significance levels for the Fisher-Yates test of significance in 2×2 contingency tables. By D. J. FINNEY

With introductory matter. Price Two Shillings and Sixpence, post free

VIII. From *Biometrika*, Vol. 35, Parts 1 and 2, pp. 191–201

Table for the calculation of working probits and weights in probit analysis. By D. J. FINNEY and W. L. STEVENS

With introductory matter. Price Two Shillings and Sixpence, post free

IX. From *Biometrika*, Vol. 36, Parts 3 and 4, pp. 267–289

Tables of autoregressive series. By M. G. KENDALL

With introductory matter. Price Two Shillings and Sixpence, post free

X. From *Biometrika*, Vol. 36, Parts 3 and 4, pp. 431–449

Tables of symmetric functions, Part I. By F. N. DAVID and M. G. KENDALL

With introductory matter. Price Two Shillings and Sixpence, post free

NEW STASTISTICAL TABLES: continued

XI. From *Biometrika*, Vol. 35, Parts 1 and 2, pp. 118–144

The distribution of the extreme deviate from the sample mean and its "studentized" form. By K. R. NAIR

Price Two Shillings and Sixpence, post free

XII. From *Biometrika*, Vol. 37, pp. 168–172 and pp. 313–325

(1) Table of the probability integral of the t-distribution

(2) Table of the χ^2 integral, and of the cumulative Poisson distribution. By H. O. HARTLEY and E. S. PEARSON

Stitched together with introductory matter. Price Five Shillings, post free

XIII. From *Biometrika*, Vol. 38, Parts 1 and 2, pp. 112–130

Charts of the power function for analysis of variance tests, derived from the non-central F-distribution. By E. S. PEARSON and H. O. HARTLEY

With introductory matter. Price Two Shillings and Sixpence, post free

XIV. From *Biometrika*, Vol. 38, Parts 3 and 4, pp. 435–462

Tables of symmetric functions. Parts II and III. By F. N. DAVID and M. G. KENDALL

With introductory matter. Price Four Shillings, post free

XV. From *Biometrika*, Vol. 38, Parts 3 and 4, pp. 423–426

A chart for the incomplete beta-function and the cumulative binomial distribution. By H. O. HARTLEY and E. R. FITCH

With introductory matter and ruler scale. Price Two Shillings and Sixpence, post free

NEW PUBLICATIONS

LOGARITHMETICA BRITANNICA

A standard Table of Logarithms to Twenty Decimal Places. By A. J. THOMPSON, Ph.D. (commenced in 1922 to commemorate the tercentenary of the publication of HENRY BRIGGS's *Arithmetica Logarithmica*).

Part II, containing the logarithms of numbers 20,000–30,000, which constitutes the ninth and final section of the Table, was finished some years ago but publication was postponed until the author's General Introduction to the complete work had been written, revised and set up in type.

This Introduction deals with the methods of interpolation required and describes in some detail the mode of construction of the table and the steps taken to safeguard the accuracy of printing. Much of the material is of general interest to table makers and table users, being the fruit of Dr Thompson's long experience gained in the production of this great standard table. At the end of the Introduction are included as Appendices:

(i) a Life of Henry Briggs by Thomas Smith, 1707, translated from the Latin by J. T. Foxell;
(ii) a table of the errors in Briggs's *Arithmetica Logarithmica* of 1624;
(iii) certain Auxiliary Tables.

PRICE: Part II (100 pp.) and Introductory matter (98 pp.) 45 shillings.

After publication, which it is expected will occur this autumn, steps will be taken to make available in two bound volumes, the nine separate sections of the table (logarithms of numbers 10,000–100,000) together with the General Introduction.

ISSUED BY THE CAMBRIDGE UNIVERSITY PRESS

on behalf of the

DEPARTMENT OF STATISTICS, UNIVERSITY COLLEGE, LONDON

BIOMETRIKA PUBLICATIONS

Issued by the
Cambridge University Press, Bentley House, London, N.W. 1
and obtainable from any bookseller

The Chances of Death and Other Studies in Evolution. Vols. I and II

BY KARL PEARSON, F.R.S.

Price 30s. net

The Life, Letters and Labours of Francis Galton Vols. I, II, IIIA, and IIIB

BY KARL PEARSON, F.R.S.

Price £3. 3s. net

Karl Pearson: An Appreciation of Some Aspects of his Life and Work

BY E. S. PEARSON

Price 10s. 6d. net

A Bibliography of the Statistical and Other Writings of Karl Pearson

COMPILED BY G. M. MORANT, WITH THE ASSISTANCE OF B. L. WELCH

Price 6s. net

"Student's" Collected Papers

EDITED BY E. S. PEARSON AND JOHN WISHART
WITH A FOREWORD BY LAUNCE McMULLEN

Price 15s. net

Karl Pearson's Early Statistical Papers

Reprinted by photo-lithography for the Biometrika Trust, with the permission of the original publishers. The Volume contains eleven papers, including the more important of the memoirs entitled "Mathematical Contributions to the Theory of Evolution", first published in the *Philosophical Transactions* of the Royal Society. The original paper deriving the χ^2-distribution, published in 1900 in the *Philosophical Magazine*, is also included.

Price 21s. net

JOURNAL
OF THE
ROYAL STATISTICAL SOCIETY

THE JOURNAL OF THE ROYAL STATISTICAL SOCIETY is published in two series: SERIES A (GENERAL), four times a year, 15*s.* each part, annual subscription £3. 1*s.* post free; SERIES B (METHODOLOGICAL), two issues a year, 22*s.* 6*d.* each part, annual subscription 45*s.* 6*d.* post free.

CONTENTS OF FORTHCOMING ISSUES

SERIES A (GENERAL), VOL. CXV, PART I, 1951

The Econometrics of Family Budgets. By H. S. HOUTHAKKER. (With Discussion.)

The Interdependence of the British Economy. By T. BARNA. (With Discussion.)

Index Numbers of the Real Product of the United Kingdom. By C. F. CARTER.

The Teaching of Statistics in Schools. A Report of the Council.

Obituary: G. UDNY YULE, C.B.E., F.R.S.

Reviews of Books, Statistical and Current Notes, Periodical Returns, Additions to Library.

SERIES B (METHODOLOGICAL), VOL. XIII, NO. 2, 1951

Some Problems in the Theory of Queues. By DAVID G. KENDALL. (With Discussion.)

The Theory of Position Finding. By H. E. DANIELS. (With Discussion).

A Review of the Literature of Systematic Sampling. By W. R. BUCKLAND.

On Two-Stage Sampling. By V. P. GODAMBE.

On Certain Probability Distributions arising from Points on a Line. By B. V. SUKHATME.

The Estimation of Standard Error from Successive Finite Differences. By P. G. GUEST.

The Interpretation of Interaction in Contingency Tables. By E. H. SIMPSON.

Complex Contingency Tables Treated by the Partition of χ^2. By H. O. LANCASTER.

The Variance of Least-Square Estimates under Linear Restraints. By S. ROSENBAUM.

Change-over Trials. By H. D. PATTERSON.

A General Technique for the Analysis of Experiments with Incorrectly Treated Plots. By P. M. GRUNDY.

Subjective Judgment in Statistical Analysis: An Experimental Study. By D. J. FINNEY.

The Expression of the Complementary Outputs of Two Products in Terms of a Common Unit of Production Effort. By G. H. JOWETT.

Inversions and Rank Correlation Coefficients. By J. DURBIN AND A. STUART.

Note on Durbin and Stuart's Formula for $E(r_s)$. By H. E. DANIELS.

ROYAL STATISTICAL SOCIETY
4 Portugal Street, London, W.C. 2

Annals of Eugenics

A JOURNAL OF HUMAN GENETICS

EDITED BY L. S. PENROSE

Vol. XVI, Pt. IV *May 1952*

Further data on genetics of microcythaemia or thalassaemia minor and Cooley's disease or thalassaemia major. By I. BIANCO, G. MONTALENTI, E. SILVESTRONI and M. SINISCALCO.

Body weight at different ages and heights. By W. F. KEMSLEY.

Blood groups in Jews from the Yemen. By A. BRZEZINSKI, J. GUREVITCH, D. HERMONI and G. MUNDEL.

A statistical study of human twinning. By NORMA MCARTHUR.

Blood agglutinogens of the Mexicans. By C. ARTEAGA, M. S. MALLÈN, A. V. BROZCO and ELENA L. UGALDO.

Simple tests for bimodality and bitangentiality. By J. B. S. HALDANE.

Birth weight and length of gestation of twins, together with maternal age, parity, and survival. By MARY N. KARN.

Data on the genetics of birth weight. By L. S. PENROSE.

Subscription price 57*s.* net per volume of four quarterly parts. Single issues 15*s.* (postage extra)

CAMBRIDGE UNIVERSITY PRESS
BENTLEY HOUSE, 200 EUSTON ROAD, LONDON, N.W. 1

The Annals of Mathematical Statistics

Issued Quarterly by the Institute of Mathematical Statistics

VOL. 23, NO. 1 CONTENTS MARCH, 1952

Abraham Wald, 1902–1950. J. WOLFOWITZ.

The Formative Years of Abraham Wald and His Work in Geometry. KARL MENGER.

Abraham Wald's Contributions to Econometrics. G. TINTNER.

The Publications of Abraham Wald.

On the Power Function of Tests of Randomness Based on Runs Up and Down. HOWARD LEVENS.

On the Structure of Balanced Incomplete Block Designs. W. S. CONNOR, JR.

Formulas for the Group Sequential Sampling of Attributes. HOWARD L. JONES.

An Application of Information Theory to Multivariate Analysis. S. KULLBACK.

Corrections for Non-normality in the Use of Two-Sample *t*- and *F*-Tests at High Significance Levels. RALPH ALLAN BRADLEY.

A Bayes Approach to a Quality Control Model. M. A. GIRSHICK and HERMAN RUBIN.

Testing a Straggler Mean in a Two-way Classification Using the Range. JACK MOSHMAN.

Note on Wilcoxon's Two-Sample Test when Ties Are Present. J. HEMELRIJK.

Subscription rate $10.00 *per year inside the Western Hemisphere and* $5.00 *per year outside the Western Hemisphere*

INQUIRIES AND SUBSCRIPTION ORDERS SHOULD BE SENT TO

CARL H. FISCHER, SECRETARY, THE INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF MICHIGAN, ANN ARBOR, MICHIGAN

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

Contents for Vol. 20, April 1952, Include:

A. CHARNES, W. W. COOPER AND B. MELLON: Blending Aviation Gasolines—A Study in Programming Interdependent Activities in an Integrated Oil Company.
A. CHARNES: Optimality and Degeneracy in Linear Programming.
M. J. FARRELL: Irreversible Demand Functions.
A. DVORETZKY, J. KIEFER AND J. WOLFOWITZ: The Inventory Problem: I. Case of Known Distributions of Demand.
MICHIO MORISHIMA: Consumer's Behavior and Liquidity Preference.
HERBERT A. SIMON: On the Application of Servomechanism Theory in the Study of Production Control.
FRANK E. BOTHWELL: The Method of Equivalent Linearization.
MARTIN SHUBIK: A Business Cycle Model with Organized Labor Considered.
GERARD DEBREU: Definite and Semidefinite Quadratic Forms.
M. HATANAKA: Note on Consolidation Within a Leontief System.
ABRAHAM WALD: On a Relation Between Changes in Demand and Price Changes.
REPORT OF THE LOUVAIN MEETING.
REPORT OF THE NEW DELHI AND PATNA MEETINGS. BOOK REVIEWS.

Published Quarterly *Subscription rate available on request*

The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics.

Subscriptions to *Econometrica* and inquiries about the work of the Society and the procedure in applying for membership should be addressed to

WILLIAM B. SIMPSON, *Secretary*
THE ECONOMETRIC SOCIETY, THE UNIVERSITY OF CHICAGO
CHICAGO 37, ILLINOIS, U.S.A.

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

EDITED BY P. C. MAHALANOBIS

VOL. 11 CONTENTS PART 2

The estimation of parameters in certain stochastic processes. By HENRY B. MANN
Statistical inference applied to classificatory problems. By C. RADHAKRISHNA RAO
Multivariate binomial and Poisson distributions. By A. S. KRISHNAMOORTHY
On errors of estimates in various types of double sampling procedure. By K. C. SEAL
Estimation of parameters from incomplete data with application to design of sample surveys. By ABRAHAM MATTHAI
Confluent hypergeometric function. By PRAN NATH
A study of recent trend in infantile mortality rates in Calcutta by longitudinal survey. By K. N. MITRA, BIMALA BHATTACHARYA, KAMALA DEY, C. S. DAWN, MERCIA OBADIAH AND A. K. GAYEN
On the non-existence of certain difference sets for incomplete group designs. By S. S. SHRIKHANDE
On the non-existence of affine resolvable balanced incomplete block designs. By S. S. SHRIKHANDE
A note on the power of the best critical region for increasing sample size. By D. BASU
Some further results on errors in double sampling technique. By CHAMELI BOSE
A note on price-wage variations in cottage and factory economy. By G. C. MANDAL
BOOK REVIEWS

Annual Subscription: 30 rupees

INQUIRIES AND ORDERS MAY BE ADDRESSED TO
THE EDITOR OF SANKHYĀ
PRESIDENCY COLLEGE, CALCUTTA, INDIA

JOURNAL OF THE
AMERICAN STATISTICAL ASSOCIATION JUNE 1952
1108 16th Street, N.W., Washington 6, D.C. Vol. 47, No. 258

Estimation for Sub-Sampling Designs Employing the County as a Primary Sampling Unit — Emil H. Jebe

Some Applications of Statistics for Auditing — John Neter

Latent Structure Analysis and its Relation to Factor Analysis — Bert F. Green, Jr.

Fertility Trends and Differentials in the U.S. — Clyde V. Kiser

Book Reviews

ASA

The American Statistical Association invites as members all persons interested in:

1. Development of new theory and method.
2. Improvement of basic statistical data.
3. Application of statistical methods to practical problems.

ACCEPTANCE SAMPLING

A symposium—$1.50

ACCEPTANCE SAMPLING BY ATTRIBUTES

Developments Prior to 1941 — Paul Peach

Wartime Developments — E. G. Olds

ACCEPTANCE SAMPLING BY VARIABLES

Acceptance Sampling by Variables, with Special Reference to the Case in which Quality is Measured by Average or Dispersion — J. H. Curtiss

Use of Variables in Acceptance Inspection for Present Defective — W. Allen Wallis

Chairman's Closure — J. W. Tukey

AMERICAN STATISTICAL ASSOCIATION
1603 K Street, N.W., Washington 6, D.C.

VOLUME 39, PARTS 1 AND 2 APRIL 1952

MOMENT COEFFICIENTS OF THE k-STATISTICS IN SAMPLES FROM A FINITE POPULATION

BY JOHN WISHART

Statistical Laboratory, University of Cambridge

1. Tukey (1950) has shown how it is possible to simplify the presentation of the moment coefficients of the distribution of certain of the k-statistics in samples from a finite population, illustrating in particular by calculating the variances of k_1 and k_2, and their covariance. His method is to introduce a new sample statistic $k_{rs\ldots}$, and to work out once and for all certain non-linear functions of these as linear functions of themselves. This is much of the labour of calculating the sampling moment coefficients; the subsequent work consists of selecting those that are required and putting them together.

It should be mentioned that Dressel (1940), following the work of Dwyer (1938), introduced statistics $L_{rs\ldots}(r, s, \ldots \neq 1)$ as unbiased estimates of products $\lambda_r\lambda_s\ldots$ of the Thiele seminvariants. These are the same quantities as Tukey's $k_{rs\ldots}$ although, because of the stated limitation to the values of $r, s, \ldots$, they do not form a complete set.

It is of some interest to carry the calculations further and derive new results. In this paper the formulae are extended to the 6th order, and all moment coefficients evaluated to this order. This will carry us a good way beyond the results of Irwin & Kendall (1944). Finally, certain basic formulae are given for the 7th and 8th orders.

2. Using the notation of David & Kendall (1949) for the symmetric functions of the n sample observations, we may write $k_{rs\ldots}$ in general as

$$k_{rs\ldots} = \sum \frac{(-)^{\Sigma(\rho-1)}\Pi(\rho-1)!}{\Pi\{\pi_1!\,\pi_2!\ldots\}} \frac{r!\,s!\ldots}{\Pi\{(p_1!)^{\pi_1}(p_2!)^{\pi_2}\ldots\}} \frac{[\Pi(p_1^{\pi_1}p_2^{\pi_2}\ldots)]}{n^{(\Sigma\rho)}}, \tag{2·1}$$

where, for the first suffix of k, namely, r, we have a partition $(p_1^{\pi_1}p_2^{\pi_2}\ldots)$ of the integer r such that $\Sigma(p\pi) = r$, and having ρ parts, so that $\Sigma(\pi) = \rho$. There are similar partitions of the suffixes $s, t, \ldots$, and inside the large summation sign the symbols Π and Σ are used to denote multiplications and sums to cover all of $r, s, t, \ldots$, while the large outside summation sign denotes that we must sum over all partitions. This definition includes Fisher's k_r as particular cases.

To illustrate let us work out k_{32}. The partitions of 3 are 3, 21 and 1^3, and of 2 they are 2 and 1^2. We shall therefore have terms in [32], [31^2], [2^21], [21^3] and [1^5]. The first term is

$$\frac{(-)^{0+0}0!.0!}{1!.1!}\frac{3!.2!}{(3!)^1.(2!)^1}\frac{[3.2]}{n^{(1+1)}} = \frac{[32]}{n^{(2)}}.$$

This is followed by

$$\frac{(-)^{0+1}0!.1!}{1!.2!}\frac{3!.2!}{(3!)^1.(1!)^2}\frac{[3.1^2]}{n^{(1+2)}} = -\frac{[31^2]}{n^{(3)}},$$

and by

$$\frac{(-)^{1+0}1!.0!}{1!\,1!.1!}\frac{3!.2!}{(2!)^1(1!)^1.(2!)^1}\frac{[21.2]}{n^{(2+1)}} = -\frac{3[2^21]}{n^{(3)}}.$$

The fourth term, however, consists of two parts, namely,

$$\left.\begin{aligned}\frac{(-)^{2+0}2!.0!}{3!.1!}\frac{3!.2!}{(1!)^3.2!}\frac{[1^3.2]}{n^{(3+1)}} &= \frac{2[21^3]}{n^{(4)}}\\ \text{and}\quad \frac{(-)^{1+1}1!.1!}{1!\,1!.2!}\frac{3!.2!}{2!\,1!.(1!)^2}\frac{[21.1^2]}{n^{(2+2)}} &= \frac{3[21^3]}{n^{(4)}}\end{aligned}\right\} = \frac{5[21^3]}{n^{(4)}},$$

while the last term is $$\frac{(-)^{2+1}\,2!.1!}{3!.2!}\,\frac{3!.2!}{(1!)^3.(1!)^2}\,\frac{[1^3.1^2]}{n^{(3+2)}} = -\frac{2[1^5]}{n^{(5)}}.$$

With practice such formulae can readily be developed by combinatorial methods. Inversely, we can determine the augmented symmetric functions in terms of the k's, the general formula being

$$\frac{[rs\ldots]}{n^{(r+s+\ldots)}} = \Sigma\,\frac{1}{\Pi\{\pi_1!\,\pi_2!\ldots\}}\,\frac{r!\,s!\ldots}{\Pi\{(p_1!)^{\pi_1}(p_2!)^{\pi_2}\ldots\}}\,k_{p_1 p_1\ldots(\pi_1 \text{ factors}) p_2 p_2\ldots(\pi_2 \text{ factors})\ldots}. \qquad (2{\cdot}2)$$

In the tables below, both sets of formulae are given. To find a k in terms of augmented symmetrics read by columns from the top down to and including the diagonal, the unit entry in which is shown in bold italics. To find an augmented symmetric in terms of k's read horizontally up to and including the italicized diagonal.

3. Conversion table

1st order

$$k_1 = [1]/n$$

2nd order

	k_{11}	k_2
$[1^2]/n^{(2)}$	***1***	-1
$[2]/n$	1	***1***

3rd order

	k_{111}	k_{21}	k_3
$[1^3]/n^{(3)}$	***1***	-1	2
$[21]/n^{(2)}$	1	***1***	-3
$[3]/n$	1	3	***1***

4th order

	k_{1111}	k_{211}	k_{22}	k_{31}	k_4
$[1^4]/n^{(4)}$	***1***	-1	1	2	-6
$[21^2]/n^{(3)}$	1	***1***	-2	-3	12
$[2^2]/n^{(2)}$	1	2	***1***	.	-3
$[31]/n^{(2)}$	1	3	.	***1***	-4
$[4]/n$	1	6	3	4	***1***

5th order

	k_{11111}	k_{2111}	k_{221}	k_{311}	k_{32}	k_{41}	k_5
$[1^5]/n^{(5)}$	***1***	-1	1	2	-2	-6	24
$[21^3]/n^{(4)}$	1	***1***	-2	-3	5	12	-60
$[2^21]/n^{(3)}$	1	2	***1***	.	-3	-3	30
$[31^2]/n^{(3)}$	1	3	.	***1***	-1	-4	20
$[32]/n^{(2)}$	1	4	3	1	***1***	.	-10
$[41]/n^{(2)}$	1	6	3	4	.	***1***	-5
$[5]/n$	1	10	15	10	10	5	***1***

6th order

	k_{111111}	k_{21111}	k_{3111}	k_{2211}	k_{321}	k_{411}	k_{33}	k_{42}	k_{51}	k_6
$[1^6]/n^{(6)}$	*1*	−1	1	2	−1	−2	−6	4	6	24	−120
$[21^4]/n^{(5)}$	1	*1*	−2	−3	3	5	12	−12	−18	−60	360
$[2^21^2]/n^{(4)}$	1	2	*1*	.	−3	−3	−3	9	15	30	−270
$[31^3]/n^{(4)}$	1	3	.	*1*	.	−1	−4	4	4	20	−120
$[2^3]/n^{(3)}$	1	3	3	.	*1*	.	.	.	−3	.	30
$[321]/n^{(3)}$	1	4	3	1	.	*1*	.	−6	−4	−10	120
$[41^2]/n^{(3)}$	1	6	3	4	.	.	*1*	.	−1	−5	30
$[3^2]/n^{(2)}$	1	6	9	2	.	6	.	*1*	.	.	−10
$[42]/n^{(2)}$	1	7	9	4	3	4	1	.	*1*	.	−15
$[51]/n^{(2)}$	1	10	15	10	.	10	5	.	.	*1*	−6
$[6]/n$	1	15	45	20	15	60	15	10	15	6	*1*

4. If we define as $K_{rst\ldots}$ that same function of the N members of the population as $k_{rst\ldots}$ is of the n members of the sample, then, as shown by Tukey (1950) (who used $k'_{rst\ldots}$):

$$E_N(k_{rst\ldots}) = K_{rst\ldots}, \tag{4·1}$$

where E_N denotes the mean value for the finite population. When N becomes infinite $K_{rst\ldots}$ becomes $\kappa_r\kappa_s\kappa_t\ldots$. Consider now the problem of determining the sampling moment coefficients of the k's. To fix ideas, suppose we require the third moment coefficient of k_1 about its mean, which we shall call $M(1^3)$, following Fisher's notation (Fisher, 1929), but using M instead of μ to denote that it is the finite population result. Now

$$\begin{aligned} M(1^3) &= E_N(k_1 - K_1)^3 \\ &= E_N(k_1^3 - 3k_1^2K_1 + 3k_1K_1^2 - K_1^3). \end{aligned}$$

To use (4·1) we require k_1^2 and k_1^3 expressed as linear functions of the $k_{rst\ldots}$, and the result is then simplified by dealing in a similar manner with the resulting non-linear functions of K's. Our next problem, then, is to determine the values of powers and products of powers of the k's. This is done algebraically by using a combination of David & Kendall's (1949) 'Tables of symmetric functions' and our tables of §3. To illustrate let us work out k_2^2.

(From §3) $\quad k_2 = [2]/n - [1^2]/n^{(2)},$

(From Tables of S.F.) $\quad = (2)/n - \{(1)^2 - (2)\}/n^{(2)}$

$\quad = \{(2) - (1)^2/n\}/(n-1).$

Therefore $\quad k_2^2 = \{n^2(2)^2 - 2n(2)(1)^2 + (1)^4\}/(n^{(2)})^2.$

(From Tables of S.F.)

[4]	[31]	[2²]	[21²]	[1⁴]	
n^2 $-2n$ 1	. $-4n$ 4	n^2 $-2n$ 3	. $-2n$ 6	. . 1	
$(n-1)^2$	$-4(n-1)$	n^2-2n+3	$-2(n-3)$	1	$\div(n^{(2)})^2$

(From §3)

$$\begin{array}{r|ccccc} & k_4 & k_{31} & k_{22} & k_{211} & k_{1111} \\ & n-1 & 4n-4 & 3n-3 & 6n-6 & n-1 \\ & & -4n+4 & . & -12n+12 & -4n+4 \\ \div n^{(2)} & & & n^2-2n+3 & 2n^2-4n+6 & n^2-2n+3 \\ & & & & -2n^2+10n-12 & -2n^2+10n-12 \\ & & & & & n^2-5n+6 \\ \hline = & n-1 & . & n(n+1) & . & . \quad \div n^{(2)} \end{array}$$

$$k_2^2 = \frac{1}{n}k_4 + \left(1+\frac{2}{n-1}\right)k_{22}.$$

A similar formula holds for K_2^2 in terms of K_4, K_{22} and N.

Any such result can be worked out by a combinatorial procedure, using Fisher's rules. The note by Kendall following this paper establishes the validity of the combinatorial method. In the present case the patterns with non-zero coefficients are obviously

2	2	4
2	2	4

2	.	2
.	2	2
2	2	4

1	1	2
1	1	2
2	2	4

with numerical coefficients 1, 1 and 2, and n-coefficients n^{-1}, 1 and $(n-1)^{-1}$. Note that a pattern occurs consisting of diagonal entries only; this is a particular case of that in which the columns may be divided into two or more classes, each confined to different sets of rows. The n-coefficient of such a pattern is the product of the n-coefficients of the sub-patterns comprising it. Fisher was able to ignore such patterns because he proceeded straight to cumulants. On the other hand, the following patterns, which correspond to the vanishing terms in the above algebraic derivation, can be seen from the rules to have zero coefficients:

2	1	3
.	1	1
2	2	4

2	.	2
.	1	1
.	1	1
2	2	4

1	1	2
1	.	1
.	1	1
2	2	4

1	.	1
1	.	1
.	1	1
.	1	1
2	2	4

In fact if the expression to be evaluated does not contain a k_1, then there can be no $k_{rs\ldots}$ on the right with a unit suffix.

If we now take the expectation of k_2^2 over the finite population we get

$$E_N(k_2^2) = \frac{1}{n}K_4 + \left(1+\frac{2}{n-1}\right)K_{22},$$

which for an infinite population becomes

$$E(k_2^2) = \mu'(2^2) = \frac{1}{n}\kappa_4 + \left(1+\frac{2}{n-1}\right)\kappa_2^2,$$

which is otherwise known to be correct. For Fisher's result is that

$$\kappa(2^2) = \frac{1}{n}\kappa_4 + \frac{2}{n-1}\kappa_2^2,$$

in addition to which we know that

$$\mu'(2^2) = \kappa(2^2) + \kappa^2(2) = \kappa(2^2) + \kappa_2^2.$$

We may take advantage of Fisher's rule for the addition of a k_1 provided we recognize that, in addition to adding 1 to existing suffices in all possible ways, and dividing by n, we must also add 1 as a new suffix without change of coefficient. Thus

$$k_2 k_1 = \frac{1}{n} k_3 + k_{21} \quad (\text{from } k_2 = k_2)$$

and

$$k_2 k_1^2 = \frac{1}{n^2} k_4 + \frac{2}{n} k_{31} + \frac{1}{n} k_{22} + k_{211}.$$

Other results may be derived from the above. Thus from the 2nd order result

$$k_1^2 = \frac{1}{n} k_2 + k_{11} \quad (\text{from } k_1 = k_1)$$

we have

$$k_2 k_{11} = k_2\left(k_1^2 - \frac{1}{n} k_2\right)$$

$$= \frac{2}{n} k_{31} - \frac{2}{n^{(2)}} k_{22} + k_{211}.$$

It should be noticed that Tukey (1950, p. 513) has an error in this result; also that he did not attempt to be comprehensive in giving formulae for all combinations of the $k_{rst\ldots}$ for the 4th order.

The full table of such relations is given in the following section, to the 6th order. They are set out in order of calculation and, for each order, a line divides the comparatively small number of relations which were worked out directly (these can be developed by combinatorial rules, including the rule for the adding of a k_1) from the remainder which were derived from those already worked out (including results of lower order). They were all carefully checked.

5. FORMULAE FOR POWERS AND PRODUCTS OF THE k'S

2nd order

$$k_1^2 = k_2/n + k_{11}$$

3rd order

$$k_2 k_1 = k_3/n + k_{21},$$
$$k_1^3 = k_3/n^2 + 3k_{21}/n + k_{111},$$
$$k_{11} k_1 = 2k_{21}/n + k_{111}.$$

4th order

$$k_3 k_1 = k_4/n + k_{31},$$
$$k_2^2 = k_4/n + (n+1)\,k_{22}/(n-1),$$
$$k_2 k_1^2 = k_4/n^2 + 2k_{31}/n + k_{22}/n + k_{211},$$
$$k_1^4 = k_4/n^3 + 4k_{31}/n^2 + 3k_{22}/n^2 + 6k_{211}/n + k_{1111}.$$

$$k_{21} k_1 = k_{31}/n + k_{22}/n + k_{211},$$
$$k_2 k_{11} = 2k_{31}/n - 2k_{22}/n^{(2)} + k_{211},$$
$$k_{11} k_1^2 = 2k_{31}/n^2 + 2k_{22}/n^2 + 5k_{211}/n + k_{1111},$$
$$k_{111} k_1 = 3k_{211}/n + k_{1111},$$
$$k_{11}^2 = 2k_{22}/n^{(2)} + 4k_{211}/n + k_{1111}.$$

5th order

$$k_4 k_1 = k_5/n + k_{41},$$
$$k_3 k_2 = k_5/n + (n+5)\,k_{32}/(n-1),$$
$$k_3 k_1^2 = k_5/n^2 + 2k_{41}/n + k_{32}/n + k_{311},$$
$$k_2^2 k_1 = k_5/n^2 + k_{41}/n + 2(n+1)\,k_{32}/n^{(2)} + (n+1)\,k_{221}/(n-1),$$
$$k_2 k_1^3 = k_5/n^3 + 3k_{41}/n^2 + 4k_{32}/n^2 + 3k_{311}/n + 3k_{221}/n + k_{2111},$$
$$k_1^5 = k_5/n^4 + 5k_{41}/n^3 + 10k_{32}/n^3 + 10k_{311}/n^2 + 15k_{221}/n^2 + 10k_{2111}/n + k_{11111}.$$

$$k_{31} k_1 = k_{41}/n + k_{32}/n + k_{311},$$
$$k_3 k_{11} = 2k_{41}/n - 6k_{32}/n^{(2)} + k_{311},$$
$$k_{22} k_1 = 2k_{32}/n + k_{221},$$
$$k_{21} k_2 = k_{41}/n + (n-3)\,k_{32}/n^{(2)} + (n+1)\,k_{221}/(n-1),$$
$$k_{21} k_1^2 = k_{41}/n^2 + 3k_{32}/n^2 + 2k_{311}/n + 3k_{221}/n + k_{2111},$$
$$k_2 k_{11} k_1 = 2k_{41}/n^2 + 2(n-3)\,k_{32}/(nn^{(2)}) + 3k_{311}/n + 2(n-2)\,k_{221}/n^{(2)} + k_{2111},$$
$$k_{11} k_1^3 = 2k_{41}/n^3 + 6k_{32}/n^3 + 7k_{311}/n^2 + 12k_{221}/n^2 + 9k_{2111}/n + k_{11111},$$
$$k_{211} k_1 = k_{311}/n + 2k_{221}/n + k_{2111},$$
$$k_{21} k_{11} = 2k_{32}/n^{(2)} + 2k_{311}/n + 2(n-2)\,k_{221}/n^{(2)} + k_{2111},$$
$$k_2 k_{111} = 3k_{311}/n - 6k_{221}/n^{(2)} + k_{2111},$$
$$k_{111} k_1^2 = 3k_{311}/n^2 + 6k_{221}/n^2 + 7k_{2111}/n + k_{11111},$$
$$k_{11}^2 k_1 = 4k_{32}/(nn^{(2)}) + 4k_{311}/n^2 + 2(5n-4)\,k_{221}/(nn^{(2)}) + 8k_{2111}/n + k_{11111},$$
$$k_{1111} k_1 = 4k_{2111}/n + k_{11111},$$
$$k_{111} k_{11} = 6k_{221}/n^{(2)} + 6k_{2111}/n + k_{11111}.$$

6th order

$$k_5 k_1 = k_6/n + k_{51},$$
$$k_4 k_2 = k_6/n + (n+7)\,k_{42}/(n-1) + 6k_{33}/(n-1),$$
$$k_3^2 = k_6/n + 9k_{42}/(n-1) + (n+8)\,k_{33}/(n-1) + 6nk_{222}/(n-1)^{(2)},$$
$$k_4 k_1^2 = k_6/n^2 + 2k_{51}/n + k_{42}/n + k_{411},$$
$$k_3 k_2 k_1 = k_6/n^2 + k_{51}/n + (n+5)\,k_{42}/n^{(2)} + (n+5)\,k_{33}/n^{(2)} + (n+5)\,k_{321}/(n-1),$$
$$k_3^2 k_1 = k_6/n^2 + 3(n+3)\,k_{42}/n^{(2)} + 4n(n-2)\,k_{33}/(n^{(2)})^2 + (n+1)(n+3)\,k_{222}/(n-1)^2,$$
$$k_3 k_1^3 = k_6/n^3 + 3k_{51}/n^2 + 3k_{42}/n^2 + k_{33}/n^2 + 3k_{411}/n + 3k_{321}/n + k_{3111},$$
$$k_2^2 k_1^2 = k_6/n^3 + 2k_{51}/n^2 + (3n+1)\,k_{42}/(nn^{(2)}) + 2(n+1)\,k_{33}/(nn^{(2)}) + k_{411}/n + 4(n+1)\,k_{321}/n^{(2)} + (n+1)\,k_{222}/n^{(2)} + (n+1)\,k_{2211}/(n-1),$$
$$k_2 k_1^4 = k_6/n^4 + 4k_{51}/n^3 + 7k_{42}/n^3 + 4k_{33}/n^3 + 6k_{411}/n^2 + 16k_{321}/n^2 + 3k_{222}/n^2 + 4k_{3111}/n + 6k_{2211}/n + k_{21111},$$
$$k_1^6 = k_6/n^5 + 6k_{51}/n^4 + 15k_{42}/n^4 + 10k_{33}/n^4 + 15k_{411}/n^3 + 60k_{321}/n^3 + 15k_{222}/n^3 + 20k_{3111}/n^2 + 45k_{2211}/n^2 + 15k_{21111}/n + k_{111111}.$$

$$k_{41} k_1 = k_{51}/n + k_{42}/n + k_{411},$$
$$k_4 k_{11} = 2k_{51}/n - 8k_{42}/n^{(2)} - 6k_{33}/n^{(2)} + k_{411},$$
$$k_{32} k_1 = k_{42}/n + k_{33}/n + k_{321},$$
$$k_{31} k_2 = k_{51}/n - 2k_{42}/n^{(2)} + k_{33}/n + (n+5)\,k_{321}/(n-1),$$
$$k_3 k_{21} = k_{51}/n + (n-4)\,k_{42}/n^{(2)} - 3k_{33}/n^{(2)} + (n+5)\,k_{321}/(n-1) - 6nk_{222}/n^{(3)},$$
$$k_{22} k_2 = 2k_{42}/n - 2k_{33}/n^{(2)} + (n+3)\,k_{222}/(n-1),$$

$$k_{31}k_1^2 = k_{51}/n^2 + 2k_{42}/n^2 + k_{33}/n^2 + 2k_{411}/n + 3k_{321}/n + k_{3111},$$

$$k_3k_{11}k_1 = 2k_{51}/n^2 + 2(n-4)\,k_{42}/(nn^{(2)}) - 6k_{33}/(nn^{(2)}) + 3k_{411}/n + 2(n-4)\,k_{321}/n^{(2)} + k_{3111},$$

$$k_{22}k_1^2 = 2k_{42}/n^2 + 2k_{33}/n^2 + 4k_{321}/n + k_{222}/n + k_{2211},$$

$$k_{21}k_2k_1 = k_{51}/n^2 + 2(n-2)\,k_{42}/(nn^{(2)}) + (n-3)\,k_{33}/(nn^{(2)}) + k_{411}/n + (3n-1)\,k_{321}/n^{(2)} + (n+1)\,k_{222}/n^{(2)} + (n+1)\,k_{2211}/(n-1),$$

$$k_2^2k_{11} = 2k_{51}/n^2 - 8k_{42}/(nn^{(2)}) + 2(n^2-2n+3)\,k_{33}/(n^{(2)})^2 + k_{411}/n + 4(n+1)\,k_{321}/n^{(2)} - 4n(n+1)\,k_{222}/(n^{(2)})^2 + (n+1)\,k_{2211}/(n-1),$$

$$k_{21}k_1^3 = k_{51}/n^3 + 4k_{42}/n^3 + 3k_{33}/n^3 + 3k_{411}/n^2 + 13k_{321}/n^2 + 3k_{222}/n^2 + 3k_{3111}/n + 6k_{2211}/n + k_{21111},$$

$$k_2k_{11}k_1^2 = 2k_{51}/n^3 + 4(n-2)\,k_{42}/(n^2n^{(2)}) + 2(n-3)\,k_{33}/(n^2n^{(2)}) + 5k_{411}/n^2 + 4(3n-5)\,k_{321}/(nn^{(2)}) + 2(n-2)\,k_{222}/(nn^{(2)}) + 4k_{3111}/n + (5n-7)\,k_{2211}/n^{(2)} + k_{21111},$$

$$k_{11}k_1^4 = 2k_{51}/n^4 + 8k_{42}/n^4 + 6k_{33}/n^4 + 9k_{411}/n^3 + 44k_{321}/n^3 + 12k_{222}/n^3 + 16k_{3111}/n^2 + 39k_{2211}/n^2 + 14k_{21111}/n + k_{111111},$$

$$k_{311}k_1 = k_{411}/n + 2k_{321}/n + k_{3111},$$

$$k_{31}k_{11} = 2k_{42}/n^{(2)} + 2k_{411}/n + 2(n-4)\,k_{321}/n^{(2)} + k_{3111},$$

$$k_{221}k_1 = 2k_{321}/n + k_{222}/n + k_{2211},$$

$$k_{22}k_{11} = 2k_{33}/n^{(2)} + 4k_{321}/n - 4k_{222}/n^{(2)} + k_{2211},$$

$$k_3k_{111} = 3k_{411}/n - 18k_{321}/n^{(2)} + 12k_{222}/n^{(3)} + k_{3111},$$

$$k_{211}k_2 = k_{411}/n + 2(n-3)\,k_{321}/n^{(2)} - 2k_{222}/n^{(2)} + (n+1)\,k_{2211}/(n-1),$$

$$k_{21}^2 = k_{42}/n^{(2)} + k_{33}/n^{(2)} + k_{411}/n + 2(n-3)\,k_{321}/n^{(2)} + (n^2-n+4)\,k_{222}/n^{(3)} + (n+1)\,k_{2211}/(n-1),$$

$$k_{211}k_1^2 = k_{411}/n^2 + 6k_{321}/n^2 + 2k_{222}/n^2 + 2k_{3111}/n + 5k_{2211}/n + k_{21111},$$

$$k_{21}k_{11}k_1 = 2k_{42}/(nn^{(2)}) + 2k_{33}/(nn^{(2)}) + 2k_{411}/n^2 + 2(5n-6)\,k_{321}/(nn^{(2)}) + 2(n-2)\,k_{222}/(nn^{(2)}) + 3k_{3111}/n + (5n-7)\,k_{2211}/n^{(2)} + k_{21111},$$

$$k_2k_{111}k_1 = 3k_{411}/n^2 + 6(n-3)\,k_{321}/(nn^{(2)}) - 6k_{222}/(nn^{(2)}) + 4k_{3111}/n + 3(n-3)\,k_{2211}/n^{(2)} + k_{21111},$$

$$k_2k_{11}^2 = 4k_{42}/(nn^{(2)}) - 4k_{33}/(n^{(2)})^2 + 4k_{411}/n^2 + 8(n-3)\,k_{321}/(nn^{(2)}) + 2(n^2-n+4)\,k_{222}/(n^{(2)})^2 + 4k_{3111}/n + 4(n-2)\,k_{2211}/n^{(2)} + k_{21111},$$

$$k_{111}k_1^3 = 3k_{411}/n^3 + 18k_{321}/n^3 + 6k_{222}/n^3 + 10k_{3111}/n^2 + 27k_{2211}/n^2 + 12k_{21111}/n + k_{111111},$$

$$k_{11}^2k_1^2 = 4k_{42}/(n^2n^{(2)}) + 4k_{33}/(n^2n^{(2)}) + 4k_{411}/n^3 + 8(4n-3)\,k_{321}/(n^2n^{(2)}) + 2(5n-4)\,k_{222}/(n^2n^{(2)}) + 12k_{3111}/n^2 + 2(17n-16)\,k_{2211}/(nn^{(2)}) + 13k_{21111}/n + k_{111111},$$

$$k_{2111}k_1 = k_{3111}/n + 3k_{2211}/n + k_{21111},$$

$$k_{211}k_{11} = 4k_{321}/n^{(2)} + 2k_{222}/n^{(2)} + 2k_{3111}/n + 2(2n-3)\,k_{2211}/n^{(2)} + k_{21111},$$

$$k_{21}k_{111} = 6k_{321}/n^{(2)} - 6k_{222}/n^{(3)} + 3k_{3111}/n + 3(n-3)\,k_{2211}/n^{(2)} + k_{21111},$$

$$k_2k_{1111} = 4k_{3111}/n - 12k_{2211}/n^{(2)} + k_{21111},$$

$$k_{1111}k_1^2 = 4k_{3111}/n^2 + 12k_{2211}/n^2 + 9k_{21111}/n + k_{111111},$$

$$k_{111}k_{11}k_1 = 12k_{321}/(nn^{(2)}) + 6k_{222}/(nn^{(2)}) + 6k_{3111}/n^2 + 6(4n-3)\,k_{2211}/(nn^{(2)}) + 11k_{21111}/n + k_{111111},$$

$$k_{11}^3 = 4k_{33}/(n^{(2)})^2 + 24k_{321}/(nn^{(2)}) + 8(n-2)\,k_{222}/(n^{(2)})^2 + 8k_{3111}/n^2 + 6(5n-4)\,k_{2211}/(nn^{(2)}) + 12k_{21111}/n + k_{111111},$$

$$k_{11111}k_1 = 5k_{21111}/n + k_{111111},$$

$$k_{1111}k_{11} = 12k_{2211}/n^{(2)} + 8k_{21111}/n + k_{111111},$$

$$k_{111}^2 = 6k_{222}/n^{(3)} + 18k_{2211}/n^{(2)} + 9k_{21111}/n + k_{111111}.$$

6. As already indicated, operating with E_N on the first sets of formulae in § 5 will give the appropriate moment coefficients about the origin, and with E the moment coefficients about the origin for the infinite population case, corresponding to Fisher's cumulants. In proceeding to find the moment coefficients about the mean, in the finite population case, much work is saved by noting that any K having a unit suffix can be dropped at the final stage. Thus

$$\begin{aligned} M(1^3) &= E_N(k_1 - K_1)^3 \\ &= E_N(k_1^3) - 3K_1 E_N(k_1^2) + 2K_1^3 \\ &= K_3/n^2 - 3K_3/(nN) + 2K_3/N^2 \\ &= \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{n} - \frac{2}{N}\right) K_3 \\ &= \alpha\left(\alpha - \frac{1}{N}\right) K_3 \end{aligned}$$

if we write $\alpha = \frac{1}{n} - \frac{1}{N}$.

Again,

$$\begin{aligned} M(21^3) &= E_N\{(k_2 - K_2)(k_1 - K_1)^3\} \\ &= E_N k_2(k_1^3 - 3k_1^2 K_1 + 3k_1 K_1^2 - K_1^3) - K_2 M(1^3), \end{aligned}$$

the first part of which is

$$\begin{aligned} &E_N(k_2 k_1^3) - 3K_1 E_N(k_2 k_1^2) + 3K_1^2 E_N(k_2 k_1) - K_2 K_1^3 \\ &= \frac{1}{n^3} K_5 + \frac{4}{n^2} K_{32} - 3K_1\left(\frac{1}{n^2} K_4 + \frac{2}{n} K_{31} + \frac{1}{n} K_{22} + K_{211}\right) + 3K_1^2\left(\frac{1}{n} K_3 + K_{21}\right) - K_2 K_1^3 \\ &= \frac{1}{n^3} K_5 + \frac{4}{n^2} K_{32} - \frac{3}{n^2 N} K_5 - \frac{6}{nN} K_{32} - \frac{6}{nN} K_{32} + \frac{3}{nN^2} K_5 + \frac{3}{nN} K_{32} + \frac{9}{N^2} K_{32} - \frac{1}{N^3} K_5 - \frac{4}{N^2} K_{32} \\ &= \left(\frac{1}{n} - \frac{1}{N}\right)^3 K_5 + \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{4}{n} - \frac{5}{N}\right) K_{32} \\ &= \alpha^3 K_5 + \left\{\alpha\left(\alpha - \frac{1}{N}\right) + 3\alpha^2\right\} K_{32}. \end{aligned}$$

It follows that $$M(21^3) = \alpha^3 K_5 - \alpha\left(\alpha - \frac{1}{N}\right)(K_3 K_2 - K_{32}) + 3\alpha^2 K_{32}.$$

The last term in this formula has obvious associations with its infinite population analogue $3\kappa_3\kappa_2/n^2$, which is the term $3\kappa(21)\kappa(1^2)$ which has to be subtracted from $\mu(21^3)$ to obtain $\kappa(21^3)$. We are not concerned here with the cumulants of the multivariate distribution in the finite population case, but it should perhaps be mentioned that if we accepted Tukey's concept of an *infinite* population of samples of size n from a population of size N (Tukey, 1950, pp. 515–16) we might define a cumulant $K(21^3)$ which would be obtained from $M(21^3)$ by subtracting $3M(21)M(1^2)$, i.e. $3\alpha^2 K_3 K_2$, a term which could be combined with the last term in the above formula. But this question really needs further consideration.

The final results of our operations are given in the next section. The simpler of the general formulae are already known; for example, Irwin & Kendall (1944) gave $M(r) = K_r$ and $M(r1) = \alpha K_{r+1}$. They also gave the equivalent, in the special case $s = 2$, of our formula for $M(rs)$.

By way of comparison it may be mentioned that Sukhatme (1944) worked out the moment coefficients about the origin of the simultaneous distribution of the m_r, the sample estimates of the finite population moment coefficients M_r. Skellam's (1949) somewhat simpler results dealt similarly with the simultaneous distribution of the sample sums of powers (equivalent to estimates of the M'_r).

7. RESULTS

The formulae listed below for general r and s have been verified as far as the 6th order and may well be true in general. Three further expressions of the 6th order have also been worked out. Note that in addition to writing α for $n^{-1}-N^{-1}$ we use $(rst\ldots)$ to denote Fisher's $\kappa(rst\ldots)$ in which a term such as $\kappa_p\kappa_q\ldots$ is written $K_{pq}\ldots$.

$$
\begin{aligned}
M(r) &= K_r, \quad M(r1) = \alpha K_{r+1},\\
M(rs) &= (rs) - (K_rK_s - K_{rs}),\\
M(r1^2) &= \alpha^2K_{r+2} - \alpha(K_rK_2 - K_{r2}),\\
M(r1^3) &= \alpha^3K_{r+3} - \alpha(\alpha - N^{-1})(K_rK_3 - K_{r3}) + 3\alpha^2K_{r+1.2},\\
M(r21) &= n\alpha\{(r21)\} - \alpha(K_{r+1}K_2 - K_{r+1.2} + K_rK_3 - K_{r3}),\\
M(r1^4) &= \alpha^4K_{r+4} - n\alpha(\alpha^3 + N^{-3})(K_rK_4 - K_{r4}) - 3\alpha^2(K_rK_{22} - K_{r22}) + 4\alpha^2(\alpha - N^{-1})K_{r+1.3}\\
&\quad + 6\alpha^3K_{r+2.2},
\end{aligned}
$$

in which the last three terms become the cumulant correction

$$4\kappa(r1)\kappa(1^3) + 6\kappa(r1^2)\kappa(1^2) = 4\kappa_{r+1}\kappa_3/n^3 + 6\kappa_{r+2}\kappa_2/n^3$$

in the limit as $N\to\infty$.

The remaining three particular formulae of the 6th order are listed below. It will be seen that a general pattern for $M(1^r)$ is emerging; the others are suggestive of what the general pattern is going to be (see also $M(2^4)$ in §8):

$$M(1^6) = n\alpha(\alpha^5 + N^{-5})K_6 + 15n\alpha^2(\alpha^3 + N^{-3})K_{42} + 10\alpha^2(\alpha - N^{-1})^2K_{33} + 15\alpha^3K_{222},$$

in which the last three terms become the cumulant correction in the limit,

$$
\begin{aligned}
M(2^3) &= (2^3) - \frac{3}{n}(K_4K_2 - K_{42}) - 3\frac{n+1}{n-1}(K_{22}K_2 - K_{222}) + 2(K_2^3 - K_{222}),\\
M(2^21^2) &= n^2\alpha^2\{(2^21^2)\} - 2\alpha^2(K_4K_2 - K_{42}) - 2\alpha(K_{22}K_2 - K_{222}) + \alpha(K_2^3 - K_{222})\\
&\quad + \alpha\{K_{42}/n + 2K_{222}/(n-1)\} + 2\alpha^2K_{33},
\end{aligned}
$$

in which the last two terms become in the limit the cumulant correction $\kappa(2^2)\kappa(1^2) + 2\kappa^2(21)$.

On proceeding to the limit as $N\to\infty$, subtracting where appropriate the cumulant correction, we see that only the first terms in all these formulae survive, and these are readily identified with the corresponding infinite population cumulants.

8. HIGHER ORDER FORMULAE FOR POWERS AND PRODUCTS OF THE k'S

It would take a great deal of space to write down all the formulae of §5 for even the 7th and 8th orders. But, as was indicated earlier, it is only necessary to work out a small number of these directly, the remainder being then determinable by algebraic substitution. Thus, for the 7th order, there are only 14 fundamental formulae, and, for the 8th order, 21. Many

of these contain k_1 or a power of k_1 so that they are readily determinable from lower order results, as previously shown. This method disposes of 11 out of the 14 fundamental formulae of the 7th order, and of 15 out of the 21 of the 8th order. 6 out of the 11, and 7 out of the 15, are of the type $k_p k_1^q$ and can be written down at once from the appropriate rows of David & Kendall's (1949) 'Tables of symmetric functions'.

The remaining formulae, namely, 3 of the 7th order and 6 of the 8th order, have been worked out by combinatorial methods, and checked by deriving the appropriate moment coefficients about the origin from Fisher's (1929) table of formulae for the cumulants of combinations of k-statistics in the infinite population case. For example, to determine $k_3 k_2^2$ we start with $\kappa(32^2)$, and then, noting that

$$\mu'_{12} = \kappa_{12} + \kappa_{10}\kappa_{02} + 2\kappa_{11}\kappa_{01} + \kappa_{10}\kappa_{01}^2,$$

add to it the terms $\kappa(3)\kappa(2^2)$, $2\kappa(32)\kappa(2)$ and $\kappa(3)\kappa^2(2)$, all of which are given. Finally, the combinations of κ's in the answer are replaced by the appropriate $k_{rst\ldots}$.

The fundamental $14+21$ formulae only are listed below for the benefit of those who would like to extend the results of §7.

7th order

$$k_6 k_1 = k_7/n + k_{61},$$

$$k_5 k_2 = k_7/n + (n+9)\,k_{52}/(n-1) + 20k_{43}/(n-1),$$

$$k_4 k_3 = k_7/n + 12k_{52}/(n-1) + (n+29)\,k_{43}/(n-1) + 36nk_{322}/(n-1)^{(2)},$$

$$k_5 k_1^2 = k_7/n^2 + 2k_{61}/n + k_{52}/n + k_{511},$$

$$k_4 k_2 k_1 = k_7/n^2 + k_{61}/n + (n+7)\,k_{52}/n^{(2)} + (n+19)\,k_{43}/n^{(2)} + (n+7)\,k_{421}/(n-1) + 6k_{331}/(n-1),$$

$$k_3^2 k_1 = k_7/n^2 + k_{61}/n + 9k_{52}/n^{(2)} + (2n+25)\,k_{43}/n^{(2)} + 9k_{421}/(n-1) + (n+8)\,k_{331}/(n-1) + 18k_{322}/(n-1)^{(2)} + 6nk_{2221}/(n-1)^{(2)},$$

$$k_3 k_2^2 = k_7/n^2 + 2(n+7)\,k_{52}/n^{(2)} + (n^2+22n-35)\,k_{43}/\{(n-1)\,n^{(2)}\} + (n+5)(n+7)\,k_{322}/(n-1)^2,$$

$$k_4 k_1^3 = k_7/n^3 + 3k_{61}/n^2 + 3k_{52}/n^2 + k_{43}/n^2 + 3k_{511}/n + 3k_{421}/n + k_{4111},$$

$$k_3 k_2 k_1^2 = k_7/n^3 + 2k_{61}/n^2 + 2(n+2)\,k_{52}/(nn^{(2)}) + 3(n+5)\,k_{43}/(nn^{(2)}) + k_{511}/n + 2(n+5)\,k_{421}/n^{(2)} + 2(n+5)\,k_{331}/n^{(2)} + (n+5)\,k_{322}/n^{(2)} + (n+5)\,k_{3211}/(n-1),$$

$$k_2^3 k_1 = k_7/n^3 + k_{61}/n^2 + 3(n+3)\,k_{52}/(nn^{(2)}) + (3n^2+14n-25)\,k_{43}/(n^{(2)})^2 + 3(n+3)\,k_{421}/n^{(2)} + 4(n-2)\,k_{331}/\{(n-1)\,n^{(2)}\} + 3(n+1)(n+3)\,k_{322}/\{(n-1)\,n^{(2)}\} + (n+1)(n+3)\,k_{2221}/(n-1)^2,$$

$$k_3 k_1^4 = k_7/n^4 + 4k_{61}/n^3 + 6k_{52}/n^3 + 5k_{43}/n^3 + 6k_{511}/n^2 + 12k_{421}/n^2 + 4k_{331}/n^2 + 3k_{322}/n^2 + 4k_{4111}/n + 6k_{3211}/n + k_{31111},$$

$$k_2^2 k_1^3 = k_7/n^4 + 3k_{61}/n^3 + (5n-1)\,k_{52}/(n^2n^{(2)}) + (7n+5)\,k_{43}/(n^2n^{(2)}) + 3k_{511}/n^2 + 3(3n+1)\,k_{421}/(nn^{(2)}) + 6(n+1)\,k_{331}/(nn^{(2)}) + 7(n+1)\,k_{322}/(nn^{(2)}) + k_{4111}/n + 6(n+1)\,k_{3211}/n^{(2)} + 3(n+1)\,k_{2221}/n^{(2)} + (n+1)\,k_{22111}/(n-1),$$

$$k_2 k_1^5 = k_7/n^5 + 5k_{61}/n^4 + 11k_{52}/n^4 + 15k_{43}/n^4 + 10k_{511}/n^3 + 35k_{421}/n^3 + 20k_{331}/n^3 + 25k_{322}/n^3 + 10k_{4111}/n^2 + 40k_{3211}/n^2 + 15k_{2221}/n^2 + 5k_{31111}/n + 10k_{22111}/n + k_{211111},$$

$$k_1^7 = k_7/n^6 + 7k_{61}/n^5 + 21k_{52}/n^5 + 35k_{43}/n^5 + 21k_{511}/n^4 + 105k_{421}/n^4 + 70k_{331}/n^4 + 105k_{322}/n^4 + 35k_{4111}/n^3 + 210k_{3211}/n^3 + 105k_{2221}/n^3 + 35k_{31111}/n^2 + 105k_{22111}/n^2 + 21k_{211111}/n + k_{1111111}.$$

8th order

$$k_7 k_1 = k_8/n + k_{71},$$

$$k_6 k_2 = k_8/n + (n+11)\,k_{62}/(n-1) + 30k_{53}/(n-1) + 20k_{44}/(n-1),$$

$$k_5 k_3 = k_8/n + 15k_{62}/(n-1) + (n+44)\,k_{53}/(n-1) + 30k_{44}/(n-1) + 60nk_{422}/(n-1)^{(2)} + 90nk_{332}/(n-1)^{(2)},$$

$$k_4^2 = k_8/n + 16k_{62}/(n-1) + 48k_{53}/(n-1) + (n+33)\,k_{44}/(n-1) + 72nk_{422}/(n-1)^{(2)} + 144nk_{332}/(n-1)^{(2)} + 24n(n+1)\,k_{2222}/(n-1)^{(3)},$$

$$k_6 k_1^2 = k_8/n^2 + 2k_{71}/n + k_{62}/n + k_{611},$$

$$k_5 k_2 k_1 = k_8/n^2 + k_{71}/n + (n+9)\,k_{62}/n^{(2)} + (n+29)\,k_{53}/n^{(2)} + 20k_{44}/n^{(2)} + (n+9)\,k_{521}/(n-1) + 20k_{431}/(n-1),$$

$$k_4 k_3 k_1 = k_8/n^2 + k_{71}/n + 12k_{62}/n^{(2)} + (n+41)\,k_{53}/n^{(2)} + (n+29)\,k_{44}/n^{(2)} + 12k_{521}/(n-1) + (n+29)\,k_{431}/(n-1) + 36k_{422}/(n-1)^{(2)} + 72k_{332}/(n-1)^{(2)} + 36nk_{3221}/(n-1)^{(2)},$$

$$k_4 k_2^2 = k_8/n^2 + 2(n+9)\,k_{62}/n^{(2)} + 8(5n-7)\,k_{53}/\{(n-1)\,n^{(2)}\} + (n^2+26n-39)\,k_{44}/\{(n-1)\,n^{(2)}\} + (n+7)\,(n+9)\,k_{422}/(n-1)^2 + 12(n+9)\,k_{332}/(n-1)^2,$$

$$k_3^2 k_2 = k_8/n^2 + (n+20)\,k_{62}/n^{(2)} + 2(n^2+22n-32)\,k_{53}/\{(n-1)\,n^{(2)}\} + 9(3n-5)\,k_{44}/\{(n-1)\,n^{(2)}\} + 9(n^2+9n-20)\,k_{422}/\{(n-1)^2\,(n-2)\} + (n^3+17n^2+104n-320)\,k_{332}/\{(n-1)^2\,(n-2)\} + 6n(n+5)\,k_{2222}/\{(n-1)^2\,(n-2)\},$$

$$k_5 k_1^3 = k_8/n^3 + 3k_{71}/n^2 + 3k_{62}/n^2 + k_{53}/n^2 + 3k_{611}/n + 3k_{521}/n + k_{5111},$$

$$k_4 k_2 k_1^2 = k_8/n^3 + 2k_{71}/n^2 + 2(n+3)\,k_{62}/(nn^{(2)}) + 2(n+13)\,k_{53}/(nn^{(2)}) + (n+19)\,k_{44}/(nn^{(2)}) + k_{611}/n + 2(n+7)\,k_{521}/n^{(2)} + 2(n+19)\,k_{431}/n^{(2)} + (n+7)\,k_{422}/n^{(2)} + 6k_{332}/n^{(2)} + (n+7)\,k_{4211}/(n-1) + 6k_{3311}/(n-1),$$

$$k_3^2 k_1^2 = k_8/n^3 + 2k_{71}/n^2 + (n+8)\,k_{62}/(nn^{(2)}) + 2(n+17)\,k_{53}/(nn^{(2)}) + (2n+25)\,k_{44}/(nn^{(2)}) + k_{611}/n + 18k_{521}/n^{(2)} + 2(2n+25)\,k_{431}/n^{(2)} + 9k_{422}/(n-1)^{(2)} + (n^2+6n+20)\,k_{332}/n^{(3)} + 9k_{4211}/(n-1) + (n+8)\,k_{3311}/(n-1) + 36k_{3221}/(n-1)^{(2)} + 6k_{2222}/(n-1)^{(2)} + 6nk_{22211}/(n-1)^{(2)},$$

$$k_3 k_2^2 k_1 = k_8/n^3 + k_{71}/n^2 + 2(n+7)\,k_{62}/(nn^{(2)}) + 4(n^2+14n-21)\,k_{53}/(n^{(2)})^2 + (n^2+22n-35)\,k_{44}/(n^{(2)})^2 + 2(n+7)\,k_{521}/n^{(2)} + (n^2+22n-35)\,k_{431}/\{(n-1)\,n^{(2)}\} + (n+5)\,(n+7)\,k_{422}/\{(n-1)\,n^{(2)}\} + 2(n+5)\,(n+7)\,k_{332}/\{(n-1)\,n^{(2)}\} + (n+5)\,(n+7)\,k_{3221}/(n-1)^2,$$

$$k_2^4 = k_8/n^3 + 4(n+5)\,k_{62}/(nn^{(2)}) + 32(n-2)\,k_{53}/(n^{(2)})^2 + (3n^3+23n^2-63n+45)\,k_{44}/\{(n-1)\,(n^{(2)})^2\} + 6(n+3)\,(n+5)\,k_{422}/\{(n-1)\,n^{(2)}\} + 16(n-2)\,(n+5)\,k_{332}/\{(n-1)^2\,n^{(2)}\} + (n+1)\,(n+3)\,(n+5)\,k_{2222}/(n-1)^3,$$

$$k_4 k_1^4 = k_8/n^4 + 4k_{71}/n^3 + 6k_{62}/n^3 + 4k_{53}/n^3 + k_{44}/n^3 + 6k_{611}/n^2 + 12k_{521}/n^2 + 4k_{431}/n^2 + 3k_{422}/n^2 + 4k_{5111}/n + 6k_{4211}/n + k_{41111},$$

$$k_3 k_2 k_1^3 = k_8/n^4 + 3k_{71}/n^3 + 2(2n+1)\,k_{62}/(n^2n^{(2)}) + (5n+19)\,k_{53}/(n^2n^{(2)}) + 3(n+5)\,k_{44}/(n^2n^{(2)}) + 3k_{611}/n^2 + 6(n+2)\,k_{521}/(nn^{(2)}) + 9(n+5)\,k_{431}/(nn^{(2)}) + 3(n+5)\,k_{422}/(nn^{(2)}) + 4(n+5)\,k_{332}/(nn^{(2)}) + k_{5111}/n + 3(n+5)\,k_{4211}/n^{(2)} + 3(n+5)\,k_{3311}/n^{(2)} + 3(n+5)\,k_{3221}/n^{(2)} + (n+5)\,k_{32111}/(n-1),$$

$$k_2^3 k_1^2 = k_8/n^4 + 2k_{71}/n^3 + 4(n+2)\,k_{62}/(n^2n^{(2)}) + 2(3n^2+10n-17)\,k_{53}/\{n(n^{(2)})^2\} + (3n^2+14n-25)\,k_{44}/\{n(n^{(2)})^2\} + k_{611}/n^2 + 6(n+3)\,k_{521}/(nn^{(2)}) + 2(3n^2+14n-25)\,k_{431}/(n^{(2)})^2 + 6(n+3)\,k_{422}/\{(n-1)\,n^{(2)}\} + 2(3n^2+14n+5)\,k_{332}/(n^{(2)})^2 + 3(n+3)\,k_{4211}/n^{(2)} + 4(n-2)\,k_{3311}/\{(n-1)\,n^{(2)}\} + 6(n+1)\,(n+3)\,k_{3221}/\{(n-1)\,n^{(2)}\} + (n+1)\,(n+3)\,k_{2222}/\{(n-1)\,n^{(2)}\} + (n+1)\,(n+3)\,k_{22211}/(n-1)^2,$$

$$k_3k_1^5 = k_8/n^5+5k_{71}/n^4+10k_{62}/n^4+11k_{53}/n^4+5k_{44}/n^4+10k_{611}/n^3+30k_{521}/n^3+25k_{431}/n^3 + 15k_{422}/n^3+10k_{332}/n^3+10k_{5111}/n^2+30k_{4211}/n^2+10k_{3511}/n^2+15k_{3221}/n^2+5k_{41111}/n + 10k_{32111}/n+k_{311111},$$

$$k_2^2k_1^4 = k_8/n^5+4k_{71}/n^4+4(2n-1)\,k_{62}/(n^3n^{(2)})+4(3n+1)\,k_{53}/(n^3n^{(2)})+(7n+5)\,k_{44}/(n^3n^{(2)}) + 6k_{611}/n^3+4(5n-1)\,k_{521}/(n^2n^{(2)})+4(7n+5)\,k_{431}/(n^2n^{(2)})+2(8n+5)\,k_{422}/(n^2n^{(2)}) + 20(n+1)\,k_{332}/(n^2n^{(2)})+4k_{5111}/n^2+6(3n+1)\,k_{4211}/(nn^{(2)})+12(n+1)\,k_{3311}/(nn^{(2)}) + 28(n+1)\,k_{3221}/(nn^{(2)})+3(n+1)\,k_{2222}/(nn^{(2)})+k_{41111}/n+8(n+1)\,k_{32111}/n^{(2)} + 6(n+1)\,k_{22211}/n^{(2)}+(n+1)\,k_{221111}/(n-1),$$

$$k_2k_1^6 = k_8/n^6+6k_{71}/n^5+16k_{62}/n^5+26k_{53}/n^5+15k_{44}/n^5+15k_{611}/n^4+66k_{521}/n^4+90k_{451}/n^4 + 60k_{422}/n^4+70k_{332}/n^4+20k_{5111}/n^3+105k_{4211}/n^3+60k_{3311}/n^3+150k_{3221}/n^3 + 15k_{2222}/n^3+15k_{41111}/n^2+80k_{32111}/n^2+45k_{22211}/n^2+6k_{311111}/n + 15k_{221111}/n+k_{2111111},$$

$$k_1^8 = k_8/n^7+8k_{71}/n^6+28k_{62}/n^6+56k_{53}/n^6+35k_{44}/n^6+28k_{611}/n^5+168k_{521}/n^5 + 280k_{431}/n^5+210k_{422}/n^5+280k_{332}/n^5+56k_{5111}/n^4+420k_{4211}/n^4+280k_{3311}/n^4 + 840k_{3221}/n^4+105k_{2222}/n^4+70k_{41111}/n^3+560k_{32111}/n^3+420k_{22211}/n^3 + 56k_{311111}/n^2+210k_{221111}/n^2+28k_{2111111}/n+k_{11111111}.$$

As an example of what can be done with the aid of the above formulae we work out below the fourth moment of k_2.

$$\begin{aligned}
M(2^4) &= E_N(k_2-K_2)^4\\
&= E_N(k_2^4)-4K_2E_N(k_2^3)+6K_2^2E_N(k_2^2)-3K_2^4\\
&= (2^4)+\frac{4}{n^2}K_{62}+\frac{3}{n^2}K_{44}+\left(\frac{6}{n}+\frac{60}{n(n-1)}\right)K_{422}\\
&\quad+\frac{16(n-2)}{n(n-1)^2}K_{332}+\left(1+\frac{12}{n-1}+\frac{44}{(n-1)^2}\right)K_{2222}\\
&\quad-\frac{4}{n^2}K_6K_2-\frac{12(n+3)}{n(n-1)}K_{42}K_2-\frac{16(n-2)}{n(n-1)^2}K_{33}K_2\\
&\quad-\frac{4(n+1)\,(n+3)}{(n-1)^2}K_{222}K_2+\frac{6}{n}K_4K_2^2+\frac{6(n+1)}{n-1}K_{22}K_2^2-3K_2^4.
\end{aligned}$$

With some rearrangement of terms this becomes

$$\begin{aligned}
M(2^4) &= (2^4)-\frac{4}{n^2}(K_6K_2-K_{62})-\frac{12(n+3)}{n(n-1)}(K_{42}K_2-K_{422})\\
&\quad-\frac{16(n-2)}{n(n-1)^2}(K_{33}K_2-K_{332})-\frac{4(n+1)\,(n+3)}{(n-1)^2}(K_{222}K_2-K_{2222})\\
&\quad+\frac{6}{n}(K_4K_2^2-K_{422})+\frac{6(n+1)}{n-1}(K_{22}K_2^2-K_{2222})\\
&\quad-3(K_2^4-K_{2222})+3\left(\frac{K_{44}}{n^2}+\frac{4K_{422}}{n(n-1)}+\frac{4K_{2222}}{(n-1)^2}\right).
\end{aligned}$$

It will be noted that the sets of n-coefficients which are multiplied by -4, $+6$ and -3 respectively, up to the penultimate term, are those occurring in k_2^3, k_2^2 and k_2. The last term, in the limit as $N\to\infty$, becomes the cumulant correction $3\kappa^2(2^2)$.

My thanks are due to Mr D. A. East, who checked all the results of §5.

REFERENCES

DAVID, F. N. & KENDALL, M. G. (1949). *Biometrika*, **36**, 431.
DRESSEL, P. L. (1940). *Ann. Math. Statist.* **11**, 33 (44).
DWYER, P. S. (1938). *Ann. Math. Statist.* **9**, 1, 97.
FISHER, R. A. (1929).. *Proc. Lond. Math. Soc.* **30**, 199.
IRWIN, J. O. & KENDALL, M. G. (1944). *Ann. Eugen., Lond.*, **12**, 138.
SKELLAM, J. G. (1949). *J. R. Statist. Soc.* B, **11**, 291.
SUKHATME, P. V. (1944). *Sankhyā*, **6**, 363.
TUKEY, J. W. (1950). *J. Amer. Statist. Ass.* **45**, 501.

MOMENT-STATISTICS IN SAMPLES FROM A FINITE POPULATION

By M. G. KENDALL

Division of Research Techniques, London School of Economics

1. The problem of deriving the moment-statistics of sample moment-statistics is, in essence, extremely simple. With the usual notation for an augmented symmetric function of order ϖ we have, for samples from an infinite population,

$$E[p_1^{\pi_1}p_2^{\pi_2}\dots p_s^{\pi_s}] = n^{(\varpi)}(\mu'_{p_1})^{\pi_1}(\mu'_{p_2})^{\pi_2}\dots(\mu'_{p_s})^{\pi_s}, \tag{1}$$

and for samples from a finite population

$$E[p_1^{\pi_1}p_2^{\pi_2}\dots p_s^{\pi_s}]_n = [p_1^{\pi_1}p_2^{\pi_2}\dots p_s^{\pi_s}]_N n^{(\varpi)}/N^{(\varpi)}, \tag{2}$$

where the subscripts n and N in equation (2) mean that the symmetric functions relate respectively to the sample of n and the population of N members.

These two formulae contain all that is necessary to deal with the univariate case and can easily be extended to the multivariate case. The reasons that, in practice, this branch of the subject becomes algebraically complex are

(*a*) that the symmetric-function statistics with which we are usually concerned are not given in terms of augmented symmetrics and therefore have to be converted into them; and likewise the parent augmented symmetrics have to be converted back to the types we usually require;

(*b*) that powers and products of statistics, even those composed of augmented symmetrics, are not augmented symmetrics and have to be converted into them;

(*c*) that we may require the answer in terms of particular kinds of functions, particularly the cumulants, to deal more easily with special populations (e.g. the normal) or cases where the cumulants after a certain point can be neglected (e.g. the Edgeworth series);

(*d*) that we may require for conciseness and concinnity to use particular kinds of statistics and to express our results in terms of particular kinds of parameters.

2. Difficulties arising from (*a*) and (*b*) are now removed by the tables of symmetric functions. Difficulties under (*c*) were to some extent resolved by Fisher's combinatorial method of obtaining sampling cumulants, which is essentially a powerful short-cut. Since the publication of the paper by Irwin & Kendall (1944) it has been known how to derive the formulae for finite populations from those for infinite populations, this again being essentially a short-cut method of circumventing the straightforward use of equation (2). But hitherto no combinatorial method has been given for use in the finite case, and as Dr Wishart points out in the foregoing paper, the key lies in the generalized k-statistics suggested by Tukey (1950). The purpose of this note is to demonstrate the validity of the method developed by Wishart.

3. The generalized k-statistic $k_{pqr\dots}$ is most easily defined as having for its mean value the product of cumulants $\kappa_p\kappa_q\kappa_r\dots$. It then follows by the argument used by Irwin & Kendall that the mean value in samples from a finite population is $K_{pqr\dots}$.

Consider now the operators

$$S_p = \sum_{j=1}^{p} \frac{\partial^p}{\partial x_j^p} \quad (p = 1, 2, \dots), \tag{3}$$

and the operators δ_p (see, for example, Kendall, 1952, p. 275) such that

$$\left.\begin{aligned} \delta_p \mu'_r &= r^{(p)} \mu'_{r-p} \quad (r > p) \\ &= p! \qquad\quad (r = p) \\ &= 0 \qquad\quad (r < p). \end{aligned}\right\} \tag{4}$$

It is known that δ_p obliterates every cumulant except κ_p, i.e.

$$\left.\begin{aligned} \delta_p \kappa_q &= 0 \quad (q \neq p) \\ &= p! \quad (q = p). \end{aligned}\right\} \tag{5}$$

The expectation of a symmetric function after operation by a set of S's is, if a constant, the same as the operation by corresponding δ's on the expectation. Thus for any combination of operators S of total order ϖ, say

$$S = S_{p_1}^{\pi_1} S_{p_2}^{\pi_2} \dots S_{p_s}^{\pi_s},$$

we find that

$$S\, k_{q_1 q_1 (\chi_1 \text{ factors}) q_2 q_2 (\chi_2 \text{ factors}) \dots q_s q_s (\chi_s \text{ factors})} = 0, \tag{6}$$

unless $p_j = q_j$ and $\chi_j = \pi_j$ for $j = 1, \dots, s$, in which case

$$S\, k_{p_1 p_1 \dots (\pi_1 \text{ factors}) p_2 p_2 \dots (\pi_2 \text{ factors}) \dots} = (p_1!)^{\pi_1} \dots (p_s!)^{\pi_s}. \tag{7}$$

This is a generalization of a result (Kendall, 1940*a*) for the ordinary k-statistics.

If now a function of k-statistics on the left, say, is expressed as a linear function of the generalized k-statistics on the right, including a term $k_{p_1 p_1 \dots (\pi_1 \text{ factors}) \text{ etc.}}$, and we operate on both sides by S, everything is obliterated on the right except the coefficient of this statistic and on the left we have the result of operating on the function by S. The evaluation of coefficients is then the same as in the combinatorial method evolved by Fisher. It can be demonstrated exactly in the manner given by myself (see Kendall, 1940*b*, 1952, pp. 276–9). The only difference is that, as Wishart points out, we cannot use rule 3 (which enjoins ignoration of patterns which fall into separate blocks) because we are here evaluating moments, not cumulants.

4. This proof, the simplest I can find, depends on the introduction of parent cumulants, just as the Irwin-Kendall method depends notionally on formulae for the infinite population. This is a perfectly valid method of reasoning, but we may note that an alternative proof could be obtained without introducing the infinite population. Equations (6) and (7), in fact, concern only the obliterating properties of the operators S on the k-statistics and could be established directly after the manner (Kendall, 1940*a*) in which I originally proved the less general results for ordinary k-statistics.

5. It may also be noted that the operational methods developed (Kendall, 1940*c*) for proceeding to the multivariate case are applicable here for developing formulae for the finite population. It hardly seems worth while displaying the resulting formulae, which would have a very limited usefulness in the present stage of development, but the methods are ready if they are ever required.

6. It is also interesting to note that the above methods solve another outstanding problem: given a function of moments or cumulants, to write down an unbiased estimator of it. For example, the variance of the second moment about the mean is

$$\operatorname{var} m_2 = \frac{(n-1)^2}{n^2} \left(\frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1} \right), \tag{8}$$

but if we wish to use this to test significance, the expression on the right usually has to be estimated from the sample and we require an unbiased estimator. This is written down at sight as

$$\frac{(n-1)^2}{n^2}\left\{\frac{k_4}{n}+\frac{2k_{22}}{n-1}\right\},$$

and can then be easily converted to ordinary k-statistics or sample-moments by the use of Wishart's tables. In this case

$$k_{22}=\frac{n-1}{n+1}\left\{k_2^2-\frac{k_4}{n}\right\},$$

and hence the unbiased estimator of var m_2 is

$$\frac{(n-1)^2}{n^2(n+1)}\left\{\frac{n-1}{n}k_4+2k_2^2\right\}.$$

Parallel results hold for the case of sampling from a finite population.

REFERENCES

Irwin, J. O. & Kendall, M. G. (1944). *Ann. Eugen., Lond.*, **12**, 138.

Kendall, M. G. (1940*a*). *Ann. Eugen., Lond.*, **10**, 106.

Kendall, M. G. (1940*b*). *Ann. Eugen., Lond.*, **10**, 215.

Kendall, M. G. (1940*c*). *Ann. Eugen., Lond.*, **10**, 392.

Kendall, M. G. (1952). *The Advanced Theory of Statistics*, vol. **1**, 5th ed. London: Charles Griffin and Co.

Tukey, J. W. (1950). *J. Amer. Statist. Ass.* **45**, 501.

SOME EXACT TESTS IN MULTIVARIATE ANALYSIS

BY E. J. WILLIAMS

Commonwealth Scientific and Industrial Research Organization, Melbourne

Some exact significance tests for use in discriminant analysis are derived. A general method of deriving exact tests where sufficient or quasi-sufficient statistics do not exist is indicated.

1. INTRODUCTION

Suppose that a set of measurements of several variates is made on each of a number of individuals, belonging to different groups. A problem which frequently arises in the analysis of such data is to determine whether a single linear function of the variates can adequately discriminate between the groups to which the individuals belong, and to evaluate such a function. This problem is often stated as one in regression of a set of dependent variates on another set of variates, but we shall use here the terminology of discriminant functions; the results are the same whichever formulation of the problem is adopted.

We begin by recapitulating some of the results in the theory. With p variates, and $q+1$ groups, the set of population means may be represented as $q+1$ points in p-dimensional space. Generally, the points will lie in at most p or q dimensions, whichever is the less. If r is in fact the number of dimensions in which the population points lie, then the populations may be specified by a set of r variates, so that if and only if the points fall on a straight line will a single hypothetical discriminant function specify the populations. In particular, the scales and axes in the p dimensions may be so chosen that the p transformed variates are uncorrelated and of equal variance. Then, by a further rotation and translation, the axes can be chosen so that the sum of the projections of the $q+1$ points on any axis is zero, while the sum of squares of the projections of the $q+1$ points on any axis assumes a stationary value. The transformed variates corresponding to these new axes are the *canonical variates* of the set (see Hotelling (1936): Hotelling's formulation is symmetrical in the two sets of p and q variates, leading to two sets of canonical variates; in the present formulation, the second set of q canonical variates, corresponding to group differences, is implicit in the set of p).

The purposes of discriminant analysis may now be formulated as (i) to determine the minimum number of canonical variates in terms of which the data may be described (collinearity, coplanarity, etc., of the groups), and (ii) to estimate them as linear compounds of the original variates measured. We assume that the variates have a multivariate normal distribution, the same for each group, about the population means characterizing the groups. Then the discriminant functions, or canonical variates of the sample, are those transformed variates which give stationary values to the ratio of the sum of squares between groups to the total sum of squares. The ratios are found as the roots of a matrix equation, and the coefficients of the original variates in the corresponding discriminant function as the corresponding latent vector. Geometrically, the procedure of determining the sample discriminant functions is equivalent to determining the principal axes of a p-dimensional ellipsoid. As with the population, so with any sample, the number of latent roots not identically zero is min (p, q).

The sample latent roots, $\theta_1, \theta_2, \ldots, \theta_p$, are defined as the stationary values of the ratio of the sum of squares between groups to the total sum of squares. The discovery of the exact simultaneous distribution of these roots, for general values of p and q, on the null hypothesis that the groups do not differ (Fisher, 1939; Hsu, 1939; Roy, 1939) opened up the way for the development of exact significance tests. Subsequently the non-null distribution has been studied by many workers (Bartlett, 1947*a*, *b*; Roy, 1942*a*, *b*, 1946). Progress with the development of exact tests, where by exact we mean independent of the unknown population parameter, has, however, been delayed by two circumstances: first, since the methods have not been extensively applied, the development of such tests has not been a matter of great practical urgency; and secondly, the heavy mathematics entailed in the study of the distribution of the latent roots has made the theory somewhat unattractive.

Below is presented an approach by means of which tests independent of the population parameter may be derived. It is hoped to indicate lines along which the theory may profitably develop, answering the practical questions with exact tests, and at the same time avoiding some of the mathematical difficulties of the usual treatment. It is considered in particular that, while the study of the joint distribution of the latent roots, on the null hypothesis, is a necessary development from the theoretical point of view, and does lead to certain overall tests of significance, the study of the distribution of the individual roots (see Nanda, 1948) does not provide results of practical relevance.

2. General remarks on tests of significance

For the purposes of significance tests, the data are summarized in the set of p latent roots $\theta_1, \theta_2, \ldots, \theta_p$ (when $q < p$, only q of these roots are not identically zero). If the null hypothesis is not true, and the $q+1$ points representing the group means lie in a space of r dimensions, there will be r unknown parameters in the specification of the joint distribution of the latent roots. The problem of estimation, and accordingly of making tests of significance for the existence of the roots, would be solved if we could (i) find a sufficient set of r statistics for the parameters, and thus confine our attention to the r degrees of freedom provided by the set, or (ii) find a set of $(p-r)$ functionally independent statistics, distributed independently of the parameters, and confine our attention to the conditional distribution of the roots, for fixed values of these statistics, which would have r degrees of freedom and enable conditional estimates and significance tests to be made (see Fisher, 1934; Pitman, 1939*a*, *b*).

It is easy to show, from the form of the distribution in the non-null case, that no sufficient set of statistics exists. It is more difficult to determine whether or not any set of statistics distributed independently of the parameters exists, but it is considered unlikely that they do. In the absence of solutions provided by either of these alternatives, we need to develop tests framed somewhat differently from those usually considered.

When only one sample root is not identically zero the relevant tests are (i) the significance of the observed root, and (ii) the concordance of any proposed discriminant function (i.e. when $q = 1$; or multiple regression relationship when $p = 1$) with the data; or, more generally, fiducial limits for the hypothetical relationship. Each is a test of a linear hypothesis, and accordingly analysis of variance methods are applicable. Where there is more than one sample root, further aspects of the data can be tested: (iii) the significance of departures from collinearity, coplanarity, etc., or, in general, the concordance of the group means with a space of r dimensions.

The practical interpretation of the data, when differences among the groups exist, is most simple when only one discriminant function is found to account for this variation; however, any systematic study of this problem must take into account the more complex cases which may arise. We deal here in detail with the simplest cases, of one population parameter and p or $q = 2, 3$, as throwing light on the problems occurring in more complex cases, and give also a discussion of the problem for general values of p.

To indicate the lines of development leading to this approach, it may be pointed out that, while the population parameters are of theoretical importance, what are of practical interest are the population discriminant functions; and that it may prove simpler, as well as more useful, to make significance tests and determine fiducial limits for these functions than for the parameters. An analogy may be found in univariate analysis of variance. The sum of squares between groups has a non-central χ^2 distribution, the parameter of which may be estimated from the data and fiducial limits derived for it; but, in general, it is the group means which are of interest. Simultaneous fiducial limits for these may be derived by a method indicated by Fisher (1949, §64).

3. Tests of a proposed discriminant function

For the theoretical analysis we shall take as our p variates, not the original variates, but the set of p sample discriminant functions, denoting them by $x_1, x_2, \ldots, x_p$. Where two or more of the sample roots are equal (including zero roots) this definition is ambiguous; but we may then choose any set of variates, uncorrelated in the sample, which correspond to this root, without affecting the results of the analysis. We take the x in decreasing order of magnitude of the corresponding θ_i; so that $\theta_1 \geqslant \theta_2 \geqslant \ldots \geqslant \theta_p$. Since the scale of the x is arbitrary, we choose the scales such that the total sum of squares of each x is unity. The sums of squares between groups will then be $\theta_1, \theta_2, \ldots, \theta_p$.

We now consider any hypothetical discriminant function y, whose concordance with the data we wish to test. The null hypothesis then specifies that

(i) the population means of the groups fall on a straight line (that is, the differences among the groups are represented by a single parameter), and

(ii) the p-dimensional vector y specifies the direction of this line.

For the hypothetical discriminant function, the ratio of the sum of squares between groups to the total sum of squares, which will be called the discriminant ratio, and denoted by η, lies in the range θ_p, θ_1. Since, under the hypothesis to be tested, the ratio η corresponds to the population discriminant function, η will be a sufficient statistic for the population parameter, that is, the corresponding ratio for the population (see Bartlett, 1947*a*). The conditional distribution of the θ_i, given η, will therefore be independent of the unknown parameter, and this conditional distribution will be used in deriving tests of significance.

It now remains to derive the joint distribution of η and the θ_i in repeated sampling. Since the conditional distribution, in which the value of the parameter is irrelevant, is ultimately required, we may assume that the parameter is zero. In this case the directions represented by $x_1, x_2, \ldots, x_p$ in the space of the original variables will be random. The direction represented by y will therefore, when referred to the co-ordinate system of the x, be random in repeated sampling. We may take the direction ratios of y in this co-ordinate system as independent unit normal deviates $w_1, w_2, \ldots, w_p$, so that

$$y = \frac{\Sigma w_i x_i}{\sqrt{(\Sigma w_i^2)}}.$$

Then $$\eta = \frac{\Sigma w_i^2 \theta_i}{\Sigma w_i^2}.$$

A practically more useful specification of the hypothetical discriminant function is in terms of the $p-1$ sample latent roots after elimination of y, which we denote by

$$\phi_1, \phi_2, \ldots, \phi_{p-1}.$$

The p-dimensional ellipsoid $$\Sigma \frac{z_i^2}{1-\theta_i} = 1$$

has squared principal semi-diameters $1-\theta_1, 1-\theta_2, \ldots, 1-\theta_p$. The section by a $p-1$ space normal to the direction specified by $(w_1, w_2, \ldots, w_p)$ is an ellipsoid with squared principal semi-diameters

$$1-\phi_1, 1-\phi_2, \ldots, 1-\phi_{p-1}.$$

Hence the ϕ_k are given by the equation

$$\Sigma \frac{w_i^2(1-\theta_i)}{\theta_i - \phi_k} = 0.$$

We can now transform from the conditional distribution of the w_i given the θ_i, to that of the $p-1$ ϕ_k, and a new variate

$$s = \Sigma w^2.$$

We readily find $$w_i^2 = \frac{s(1-\eta)}{(1-\theta_i)} \frac{\prod_k (\theta_i - \phi_k)}{\prod_{j \neq i} (\theta_i - \theta_j)}.$$

The result may be verified in the following way:

$$\sum_i A_i \equiv \sum_i \frac{\prod_k (\theta_i - \phi_k)}{\prod_{j \neq i} (\theta_i - \theta_j)}$$

is a multilinear symmetric function of the $p-1$ variates ϕ_k; and on giving the ϕ_k in succession the p sets of values

$$\begin{array}{llll} \theta_2, & \theta_3, & \ldots, & \theta_p, \\ \theta_3, & \theta_4, & \ldots, & \theta_1, \\ \multicolumn{4}{c}{\ldots\ldots\ldots\ldots\ldots\ldots} \\ \theta_1, & \theta_2, & \ldots, & \theta_{p-1}, \end{array}$$

we find that the value of the expression is unity in each case; hence ΣA_i is identically unity. Likewise

$$\sum_i B_i \equiv \sum_i \frac{1}{\theta_i - \phi_h} \frac{\prod_k (\theta_i - \phi_k)}{\prod_{j \neq i} (\theta_i - \theta_j)}$$

is a multilinear symmetric function of the $p-2$ variates ϕ_k excluding ϕ_h. On giving these ϕ_k the sets of values

$$\theta_3, \quad \theta_4, \quad \ldots, \quad \theta_p, \text{ etc.},$$

we find that ΣB_i is identically zero.

Now $$w_i^2 = \frac{s(1-\eta)}{(1-\theta_i)} A_i,$$

so that
$$\begin{aligned}\sum_i w_i^2 &= \sum_i \frac{w_i^2(1-\eta)}{1-\eta} \\ &= \sum_i \frac{w_i^2(1-\theta_i)}{1-\eta} \\ &= s\sum_i A_i = s,\end{aligned}$$

while
$$\begin{aligned}\sum_i \frac{w_i^2(1-\theta_i)}{\theta_i-\phi_k} &= s(1-\eta)\sum_i \frac{A_i}{\theta_i-\phi_k} \\ &= s(1-\eta)\sum_i B_i \\ &= 0,\end{aligned}$$

$$\frac{\partial(w_1, w_2, \ldots, w_p)}{\partial(\phi_1, \phi_2, \ldots, \phi_{p-1}, s)} = \text{const.}\frac{w_1 w_2 \ldots w_p}{s}\left|\frac{\prod_{i\neq j}(\theta_i-\theta_j)\prod_{h\neq k}(\phi_h-\phi_k)}{\prod_i\prod_k(\theta_i-\phi_k)}\right|$$

$$= s^{\frac{1}{2}p-1}\frac{\prod(1-\theta_i)^{\frac{1}{2}(p-1)}\left|\prod_{h\neq k}(\phi_h-\phi_k)\right|}{\prod(1-\phi_k)^{\frac{1}{2}p}\left|\prod_i\prod_k(\theta_i-\phi_k)\right|^{\frac{1}{2}}}.$$

Clearly, s is distributed independently of the other variates, so that the joint distribution of the ϕ_k, for given θ_i, is

$$\text{const.}\frac{\prod(1-\theta_i)^{\frac{1}{2}(p-1)}\left|\prod_{h\neq k}(\phi_h-\phi_k)\right|}{\prod(1-\phi_k)^{\frac{1}{2}p}\,|\,\prod(\theta_i-\phi_k)\,|^{\frac{1}{2}}}\prod d\phi_k.$$

The joint distribution of the θ_i is known to be

$$\text{const.}\prod\theta_i^{\frac{1}{2}(q-p-1)}\prod(1-\theta_i)^{\frac{1}{2}(n-p-q-1)}\left|\prod_{i\neq j}(\theta_i-\theta_j)\right|\prod d\theta_i,$$

so that, finally, the joint distribution of the θ_i and ϕ_k is

$$\text{const.}\frac{\prod\theta_i^{\frac{1}{2}(q-p-1)}\prod(1-\theta_i)^{\frac{1}{2}(n-q-2)}\left|\prod_{i\neq j}(\theta_i-\theta_j)\prod_{h\neq k}(\phi_h-\phi_k)\right|\prod d\theta_i\prod d\phi_k}{\prod(1-\phi_k)^{\frac{1}{2}p}\,|\,\prod(\theta_i-\phi_k)\,|^{\frac{1}{2}}}.$$

It remains now to derive the conditional distribution, given η, which will be independent of the population parameter. Since the set of θ_i and ϕ_k define η precisely, one of them must be replaced by η; but to preserve symmetry we shall not carry out this transformation.

The distribution of η is
$$\text{const.}\,\eta^{\frac{1}{2}(q-2)}(1-\eta)^{\frac{1}{2}(n-q-2)}d\eta,$$

so that the conditional distribution of the θ_i and ϕ_k, given η, is

$$\text{const.}\frac{\prod\theta_i^{\frac{1}{2}(q-p-1)}\prod(1-\theta_i)^{\frac{1}{2}(n-q-2)}\left|\prod_{i\neq j}(\theta_i-\theta_j)\prod_{h\neq k}(\phi_h-\phi_k)\right|\prod d\theta_i\prod d\phi_k}{\eta^{\frac{1}{2}(q-2)}(1-\eta)^{\frac{1}{2}(n-q-2)}\prod(1-\phi_k)^{\frac{1}{2}p}\,|\,\prod(\theta_i-\phi_k)\,|^{\frac{1}{2}}d\eta}$$

or, since
$$(1-\eta)\prod(1-\phi_k) = \prod(1-\theta_i),$$

$$\text{const.}\frac{\prod\theta_i^{\frac{1}{2}(q-p-1)}}{\eta^{\frac{1}{2}(q-2)}}\prod(1-\phi_k)^{\frac{1}{2}(n-p-q-2)}\frac{\left|\prod_{i\neq j}(\theta_i-\theta_j)\prod_{h\neq k}(\phi_h-\phi_k)\right|\prod d\theta_i\prod d\phi_k}{|\,\prod(\theta_i-\phi_k)\,|^{\frac{1}{2}}d\eta}.$$

When p exceeds q, only q of the roots are distributed, the remaining $p-q$ roots being identically zero. It is desirable in such cases explicitly to modify the distribution given above, putting

$$\theta_{q+1} = \phi_{q+1} = \theta_{q+2} = \ldots = \phi_{p-1} = \theta_p = 0.$$

Then, for the remaining roots, the joint distribution is found to be

$$\text{const.} \frac{\prod \phi_k^{\frac{1}{2}(p-q-2)}}{\eta^{\frac{1}{2}(q-2)}} \prod (1-\phi_k)^{\frac{1}{2}(n-p-q-2)} \frac{\left| \prod_{i+j} (\theta_i - \theta_j) \prod_{h+k} (\phi_h - \phi_k) \right| \prod d\theta_i \prod d\phi_k}{\left| \prod (\theta_i - \phi_k) \right|^{\frac{1}{2}} d\eta}.$$

In particular, when $q = 2$, we have

$$\text{const.} (\phi_1\phi_2)^{\frac{1}{2}(p-4)} [(1-\phi_1)(1-\phi_2)]^{\frac{1}{2}(n-p-4)} \frac{(\theta_1-\theta_2)(\phi_1-\phi_2)\, d\theta_1 d\theta_2 d\phi_1 d\phi_2}{\sqrt{\{-\prod (\theta_i - \phi_k)\}}\, d\eta}.$$

These conditional distributions, being independent of the parameter, now provide the basis for any exact tests which are to be made.

(*a*) *Two-variate problems* ($p = 2$)

The general distribution may be used, but since the discriminant function y is specified by only one direction-cosine, it is adequately defined by η.

Then
$$\eta = \frac{w_1^2\theta_1 + w_2^2\theta_2}{w_1^2 + w_2^2},$$

and on transforming from w_1, w_2 to η, s, the conditional distribution of η is found to be

$$\frac{d\eta}{\pi \sqrt{\{(\theta_1-\eta)(\eta-\theta_2)\}}}.$$

The joint distribution of θ_1, θ_2 and η is therefore

$$\text{const.} (\theta_1\theta_2)^{\frac{1}{2}(q-3)} [(1-\theta_1)(1-\theta_2)]^{\frac{1}{2}(n-q-3)} \frac{(\theta_1-\theta_2)\, d\theta_1 d\theta_2 d\eta}{\sqrt{\{(\theta_1-\eta)(\eta-\theta_2)\}}},$$

and the conditional distribution of θ_1 and θ_2 is

$$\text{const.} \frac{(\theta_1\theta_2)^{\frac{1}{2}(q-3)} [(1-\theta_1)(1-\theta_2)]^{\frac{1}{2}(n-q-3)} (\theta_1-\theta_2)\, d\theta_1 d\theta_2}{\eta^{\frac{1}{2}(q-2)} (1-\eta)^{\frac{1}{2}(n-q-2)} \sqrt{\{(\theta_1-\eta)(\eta-\theta_2)\}}}.$$

We now make a change of variables to

$$v_1 = \frac{(\theta_1-\eta)(\eta-\theta_2)}{(1-\eta)(\eta-\theta_1\theta_2)};$$

$$v_2 = \frac{\theta_1\theta_2}{\eta}.$$

The joint distribution of v_1 and v_2 is

$$\text{const.} v_1^{-\frac{1}{2}} (1-v_1)^{\frac{1}{2}(n-q-3)} dv_1 v_2^{\frac{1}{2}(q-3)} (1-v_2)^{\frac{1}{2}(n-q-2)} dv_2.$$

It is seen that v_1 and v_2 are independently distributed, and so may be used to test independent aspects of the data. Moreover, the distribution does not involve η, so that it also gives the unrestricted distribution, when η is not fixed, for any value of the population parameter.

It is to be noted that, while v_1 and v_2 formally are symmetric functions of θ_1 and θ_2, they depend on these statistics in quite different ways; v_1 is dependent mainly on θ_1, while v_2 depends mainly on θ_2. This is because, for any acceptable hypothesis, η must lie in the neighbourhood of θ_1; consequently,

$$v_1 \sim \frac{\theta_1 - \eta}{1 - \eta},$$

$$v_2 \sim \theta_2.$$

Hence v_1 is appropriate for testing the concordance of the data with the form of the proposed discriminant function while v_2 gives a test for departure from collinearity. The latter test is a rather 'fluid' one, in that it depends on the choice of the hypothetical discriminant function and the resulting η. The justification for the use of v_2 is that it is independent of the unknown parameter, being dependent instead on a statistic which, while to some extent arbitrary, is yet determinate. Where the set of acceptable hypotheses as determined by v_1 is well defined, v_2 will often not differ greatly from θ_2; however, since θ_2 is the minimum value of v_2, we have a simple demonstration of the fact that the systematic use of θ_2 as a statistic with the distribution of v_2 to test departures from linearity tends to underestimate significance.

The criteria have independent F distributions;

$$\frac{(n-q-1)\,v_1}{1-v_1}$$

is distributed as F with 1 and $n-q-1$ degrees of freedom, while

$$\frac{(n-q)\,v_2}{(q-1)\,(1-v_2)}$$

is distributed as F with $q-1$ and $n-q$ degrees of freedom. However, it seems best to test the two aspects of the data in the manner indicated in the analysis of variance set out below, rather than to use v_1 and v_2 as they stand.

It may be remarked here that the discriminant function y will be sometimes rejected because of significant discrepancy in direction; at other times because of 'departure from linearity'. For any single discriminant function to be acceptable, it must clearly give non-significant departures for either aspect.

The criterion v_1 is not a monotonic function of η; from a zero value at $\eta = \theta_1$ it rises to a maximum value

$$\left(\frac{\sqrt{\theta_1} - \sqrt{\theta_2}}{1 - \sqrt{(\theta_1 \theta_2)}}\right)^2$$

at

$$\eta = \sqrt{(\theta_1 \theta_2)}.$$

Hence, for a sufficiently high level of significance in any example, no chosen discriminant function will be judged discordant with the data; in other words, no fiducial limits for the true discriminant function will exist. This is to be expected, since the difference between any two discriminant functions corresponds, not to a distance which can be increased indefinitely, but to an angle.

An overall test for the adequacy of the discriminant function y is provided by a method given by Bartlett (1939). The effect of the proposed discriminant function is eliminated by

covariance, and the discriminant ratio for the adjusted variates is tested. Since, in this case, y is a linear compound of x_1 and x_2, the elimination of its effect from either has the same result. The analysis of covariance of x_1, for instance, where μ is the angle between y and x_2, is:

	Sums of squares and products			Adjusted sum of squares x_1^2
	x_1^2	$x_1 y$	y^2	
Between groups	θ_1	$\theta_1 \sin\mu$	$\eta = \theta_1 \sin^2\mu + \theta_2 \cos^2\mu$	$\frac{(\theta_1+\theta_2-\theta_1\theta_2-\eta)}{1-\eta}\cos^2\mu$
Within groups	$1-\theta_1$	$(1-\theta_1)\sin\mu$	$1-\eta$	$\frac{(1-\theta_1)(1-\theta_2)}{1-\eta}\cos^2\mu$
Total	1	$\sin\mu$	1	$\cos^2\mu$

Denoting the adjusted discriminant ratio by u, we have

$$1-u = \frac{(1-\theta_1)(1-\theta_2)}{1-\eta} = (1-v_1)(1-v_2),$$

so that

$$u = v_1 + v_2 - v_1 v_2.$$

A significant value of u may result from a large value of v_1 or v_2 or both. The ratio

$$\frac{(n-q-1)\,u}{q(1-u)}$$

is distributed as F with q and $n-q-1$ degrees of freedom.

The criteria may be set out in analysis of variance form, as follows:

	Degrees of freedom	Sum of squares
Direction	1	$\frac{(\theta_1-\eta)(\eta-\theta_2)}{\eta(1-\eta)} = v_1(1-v_2)$
Linearity	$q-1$	$\frac{\theta_1\theta_2}{\eta} = v_2$
Residual	$n-q-1$	$\frac{(1-\theta_1)(1-\theta_2)}{1-\eta} = (1-v_1)(1-v_2)$
Total	$n-1$	1

This shows that the criteria for direction and linearity may each be tested against an independent estimate of residual variation.

When $q = 1, \theta_2 = 0$ and the analysis of variance reduces to the familiar form.

(b) Three groups ($q = 2$)

When there are three groups, there are two sample roots, θ_1 and θ_2, not identically zero, but there are also $p-2$ zero roots. Any hypothetical discriminant function y then lies in a space of p dimensions, the corresponding discriminant ratio η may range from zero to θ_1, and the analysis given above for $p = 2$ does not apply. Here again the general distribution may be used, but we shall adopt a simpler approach which is of interest.

We have $$\eta = \frac{w_1^2\theta_1 + w_2^2\theta_2}{s},$$

and shall write $$\zeta = \frac{w_1^2\theta_1 + w_2^2\theta_2}{w_1^2 + w_2^2},$$

so that $\theta_2 \leqslant \zeta \leqslant \theta_1$.

As before, $w_1, w_2, \ldots, w_p$ are, on the null hypothesis, independent unit normal deviates. The joint distribution of s, w_1 and w_2 is then found to be

$$\text{const. } e^{-\frac{1}{2}s}(s - w_1^2 - w_2^2)^{\frac{1}{2}(p-4)}\,ds\,dw_1\,dw_2.$$

We now transform the variates to find the joint distribution of η, ζ and s; after integration of this joint distribution with respect to s, the joint distribution of η and ζ is given as

$$\text{const. } \left(\frac{\zeta-\eta}{\zeta}\right)^{\frac{1}{2}(p-4)} \frac{d\eta\,d\zeta}{\zeta\sqrt{\{(\theta_1-\zeta)(\zeta-\theta_2)\}}}.$$

The joint distribution of θ_1, θ_2, η and ζ is therefore

$$\text{const. } (\theta_1\theta_2)^{\frac{1}{2}(p-3)}[(1-\theta_1)(1-\theta_2)]^{\frac{1}{2}(n-p-3)}(\theta_1-\theta_2)\left(\frac{\zeta-\eta}{\zeta}\right)^{\frac{1}{2}(p-4)} \frac{d\theta_1\,d\theta_2\,d\eta\,d\zeta}{\zeta\sqrt{\{(\theta_1-\zeta)(\zeta-\theta_2)\}}},$$

and the conditional distribution of θ_1, θ_2 and ζ, given η, is

$$\text{const. } (\theta_1\theta_2)^{\frac{1}{2}(p-3)} \frac{[(1-\theta_1)(1-\theta_2)]^{\frac{1}{2}(n-p-3)}}{(1-\eta)^{\frac{1}{2}(n-4)}} (\theta_1-\theta_2)\left(\frac{\zeta-\eta}{\zeta}\right)^{\frac{1}{2}(p-4)} \frac{d\theta_1\,d\theta_2\,d\zeta}{\zeta\sqrt{\{(\theta_1-\zeta)(\zeta-\theta_2)\}}}.$$

We now put $$v_1 = \frac{(\theta_1-\zeta)(\zeta-\theta_2)}{(1-\zeta)(\zeta-\theta_1\theta_2)},$$

$$v_2 = \frac{\theta_1\theta_2}{\zeta},$$

$$v_3 = \frac{\zeta-\eta}{1-\eta}.$$

The joint distribution of v_1, v_2 and v_3 is found to be

$$\text{const. } v_1^{-\frac{1}{2}}(1-v_1)^{\frac{1}{2}(n-p-3)}dv_1\,v_2^{\frac{1}{2}(p-3)}(1-v_2)^{\frac{1}{2}(n-p-2)}dv_2\,v_3^{\frac{1}{2}(p-4)}(1-v_3)^{\frac{1}{2}(n-p-2)}dv_3,$$

so that v_1, v_2 and v_3 are independently distributed. From their definition it is seen that v_1 and v_2 provide tests of direction and linearity respectively, attention being confined to the x_1, x_2 plane, and in fact correspond to the v_1 and v_2 defined for the discussion of the case $p = 2$. A test of the departure of the hypothetical discriminant function from the x_1, x_2 plane is provided by v_3. The tests cannot, however, be set out in the form of an analysis

of variance, as was possible for $p = 2$. This is because the criterion for an overall test of the adequacy of the hypothetical discriminant function, namely,

$$\frac{(1-\theta_1)(1-\theta_2)}{1-\eta} = (1-v_1)(1-v_2)(1-v_3),$$

is distributed not as a Beta-variate but as the square of a $B(n-p-1, p-1)$ variate. This is the result first discovered by Wilks (1932), and enables the 'residual' to be tested by means of the F distribution with $2(p-1)$ and $2(n-p-1)$ degrees of freedom.

We can, however, set out the tests for departures in direction in one analysis of variance, and the tests for departures from linearity in another. Departures in direction may lie in the p space, while departures from linearity are observable only in the x_1, x_2 plane.

Direction	Degrees of freedom	Sum of squares
In the x_1, x_2 plane	1	$\frac{(\theta_1-\zeta)(\zeta-\theta_2)}{(1-\eta)(\zeta-\theta_1\theta_2)} = v_1(1-v_2)$
In the $p-2$ space	$p-2$	$\frac{\zeta-\eta}{1-\eta} = v_2$
Residual	$n-p-1$	$\frac{\zeta(1-\theta_1)(1-\theta_2)}{(1-\eta)(\zeta-\theta_1\theta_2)} = (1-v_1)(1-v_2)$
Total	$n-2$	1

Linearity	Degrees of freedom	Sum of squares
Direction in the plane	1	$\frac{(\theta_1-\zeta)(\zeta-\theta_2)}{\zeta(1-\zeta)} = v_1(1-v_2)$
Linearity	$p-1$	$\frac{\theta_1\theta_2}{\zeta} = v_2$
Residual	$n-p-1$	$\frac{(1-\theta_1)(1-\theta_2)}{1-\zeta} = (1-v_1)(1-v_2)$
Total	$n-1$	1

These two analyses enable detailed tests to be made of the various departures from the hypothetical discriminant function. It must be recognized, however, that these tests are not independent.

The limiting case when n tends to infinity gives an exact analysis of variance. Such a case arises in practice in the canonical analysis of a set of variates with known covariance matrix. If, modifying the notation slightly, we put

$$n\theta_1 \to S_1, \quad n\eta \to S_w,$$
$$n\theta_2 \to S_2, \quad n\zeta \to S_r,$$

$$v_1 = \frac{(S_1 - S_r)(S_r - S_2)}{S_r},$$

$$v_2 = \frac{S_1 S_2}{S_r},$$

$$v_3 = S_r - S_w,$$

the joint distribution of v_1, v_2 and v_3 comes out as

$$\text{const.}\, e^{-\frac{1}{2}(v_1+v_2+v_3)} v_1^{-\frac{1}{2}} v_2^{\frac{1}{2}(p-3)} v_3^{\frac{1}{2}(p-4)} dv_1 dv_2 dv_3,$$

and the analysis of variance is then

	Degrees of freedom	Sum of squares
Direction in the plane	1	$\frac{(S_1 - S_r)(S_r - S_2)}{S_r} = v_1$
Direction in $p-2$ space	$p-2$	$S_r - S_w = v_3$
Linearity	$p-1$	$\frac{S_1 S_2}{S_r} = v_2$
Departures from hypothetical discriminant function	$2(p-1)$	$S_1 + S_2 - S_w = v_1 + v_2 + v_3$
Hypothetical discriminant function	2	S_w
Total	$2p$	$S_1 + S_2$

This analysis is applied to a practical example in a forthcoming paper (Williams, 1952).

It now remains to relate the results for $q = 2$ to those obtained in the analysis of the general case. To do this, we express v_1, v_2 and v_3 in terms of the θ_i and ϕ_k. We have

$$1 - \eta = \frac{(1-\theta_1)(1-\theta_2)}{(1-\phi_1)(1-\phi_2)},$$

$$w_1^2 = \frac{s(1-\eta)}{(1-\theta_1)} \frac{(\theta_1 - \phi_1)(\theta_1 - \phi_2)}{(\theta_1 - \theta_2)\theta_1},$$

$$w_2^2 = \frac{s(1-\eta)}{(1-\theta_2)} \frac{(\theta_2 - \phi_1)(\theta_2 - \phi_2)}{(\theta_2 - \theta_1)\theta_2},$$

so that

$$\frac{w_1^2 + w_2^2}{s} = \frac{(1-\eta)[\theta_1\theta_2(1-\phi_1)(1-\phi_2) - \phi_1\phi_2(1-\theta_1)(1-\theta_2)]}{\theta_1\theta_2(1-\theta_1)(1-\theta_2)}$$

$$= 1 - \frac{\phi_1\phi_2(1-\theta_1)(1-\theta_2)}{\theta_1\theta_2(1-\phi_1)(1-\phi_2)}.$$

For brevity we write

$$P = \theta_1 + \theta_2 - \phi_1 - \phi_2,$$

$$Q = \theta_1\theta_2 - \phi_1\phi_2,$$

$$R = \theta_1\theta_2(\phi_1 + \phi_2) - \phi_1\phi_2(\theta_1 + \theta_2);$$

then it can be shown that $R = (\theta_1+\theta_2)Q - \theta_1\theta_2 P$

$$= (\phi_1+\phi_2)Q - \phi_1\phi_2 P,$$

$$PR-Q^2 = -\Pi(\theta_i-\phi_k)$$

$$= (\theta_1-\phi_1)(\theta_1-\phi_2)(\phi_1-\theta_2)(\theta_2-\phi_2).$$

We find that $\zeta = \dfrac{s\eta}{w_1^2+w_2^2}$

$$= \theta_1\theta_2\frac{P-Q}{Q-R},$$

$$1-\zeta = (1-\theta_1)(1-\theta_2)\frac{Q}{Q-R},$$

$$v_2 = \frac{Q-R}{P-Q},$$

$$1-v_2 = \frac{P-2Q+R}{P-Q},$$

$$v_1 = \frac{PR-Q^2}{Q(P-2Q+R)},$$

$$v_1(1-v_2) = \frac{PR-Q^2}{Q(P-Q)}$$

$$= \frac{-\Pi(\theta_i-\phi_k)}{Q(P-Q)},$$

$$(1-v_1)(1-v_2) = \frac{Q-R}{Q},$$

$$v_3 = \phi_1\phi_2\frac{P-Q}{Q-R},$$

$$1-v_3 = (1-\phi_1)(1-\phi_2)\frac{Q}{Q-R},$$

$$v_1(1-v_3) = (1-\phi_1)(1-\phi_2)\frac{PR-Q^2}{(Q-R)(P-2Q+R)},$$

$$(1-v_1)(1-v_3) = (1-\phi_1)(1-\phi_2)\frac{P-Q}{P-2Q+R}.$$

(c) *Three or more variates; more than three groups* ($p, q \geqslant 3$)

When p and q are greater than 2, the joint distribution of the latent roots is more complicated. A method of analysis to give tests for direction and linearity is given by Bartlett (1951).

For an overall test of significance, the method of Bartlett already referred to may be used. On the hypothesis to be tested, the product of the within-group ratios, after adjustment

by covariance with the hypothetical discriminant function, will be distributed independently of the parameter. The test function thus derived is

$$1-u = \frac{\Pi(1-\theta_i)}{1-\eta} = \Pi(1-\phi_k),$$

which has already appeared in the simpler cases discussed above. When $p = 2$ it has a Beta-distribution, and when $p = 3$ or $q = 2$ (that is, when it is the product of two adjusted ratios) it is distributed as the square of a Beta-variate. For $p = 3$ the distribution is

$$\text{const. } v^{q-1}(1-v)^{n-q-3}\,dv,$$

where

$$(1-v)^2 = (1-u),$$

so that

$$\frac{(n-q-2)\,v}{q(1-v)}$$

has the F distribution with $2q$ and $2(n-q-2)$ degrees of freedom. When $q = 2$ the distribution is

$$\text{const. } v^{p-2}(1-v)^{n-p-2}\,dv,$$

so that

$$\frac{(n-p-1)\,v}{(p-1)(1-v)}$$

has the F distribution with $2(p-1)$ and $2(n-p-1)$ degrees of freedom.

For larger values of p and q the exact analytical form of the distribution is not known, but Bartlett (1938) has developed approximations using the χ^2 distribution, which are satisfactory in most cases for making significance tests (see also Hsu, 1940). A more serious criticism of the criterion is that, since the same value of η corresponds to widely differing forms of the hypothetical discriminant function, it is difficult to interpret a significant result given by the overall test.

4. NUMERICAL EXAMPLE

To illustrate the use of the tests outlined above, we apply them to some data presented by Bartlett (1947*b*, pp. 177–9). This is an experiment on the effects of eight treatments, each replicated eight times, on yield of straw (x_1 in Bartlett's notation) and of grain (x_2). We shall not question the relevance of discriminant analysis in this example, though it seems that for practical purposes some linear combination of x_1 and x_2 representing total value, rather than one maximizing the discriminant ratio, will be the appropriate one to study.

After eliminating block differences, we have

$$n = 56, \quad p = 2, \quad q = 7,$$

and Bartlett gives (in our notation) $\theta_1 = 0{\cdot}47698,$

$$\theta_2 = 0{\cdot}05934,$$

the discriminant function corresponding to θ_1 being

$$x_2 - 0{\cdot}535x_1.$$

It is reasonable to inquire first of all whether x_2 alone is a satisfactory discriminant function (in other words, whether the coefficient of x_1 differs significantly from zero). For x_2 we have

$$\eta = 0{\cdot}31570,$$

so that the sums of squares for the analysis of variance are as follows:

$$\text{Direction: } \frac{0{\cdot}16128 \times 0{\cdot}25636}{0{\cdot}31570 \times 0{\cdot}68430} = 0{\cdot}19139,$$

$$\text{Linearity: } \frac{0{\cdot}47698 \times 0{\cdot}05934}{0{\cdot}31570} = 0{\cdot}08965,$$

$$\text{Residual: } \frac{0{\cdot}52302 \times 0{\cdot}94066}{0{\cdot}68430} = 0{\cdot}71896,$$

and the analysis of variance is

	Degrees of freedom	Sum of squares	Mean square
Direction	1	0·19139	0·19139
Linearity	6	0·08965	0·01494
Residual	48	0·71896	0·01498
Total	55	1·00000	

The data are discordant with the proposed discriminant function x_2, with significance at the 1 % level, so that x_2 alone is unsatisfactory.

We next inquire what set of discriminant functions is acceptable in direction, assuming no departures from linearity. If the tabular value of the F distribution for 1 and 48 degrees of freedom at the required significance level is F, we have for the limiting value of η,

$$\frac{48(\theta_1 - \eta)(\eta - \theta_2)}{\eta(1-\theta_1)(1-\theta_2)} = F.$$

At the 5 % level, $F = 4{\cdot}0427$, giving $\eta = 0{\cdot}42890$; at the 1 % level, $F = 7{\cdot}1942$, giving $\eta = 0{\cdot}39002$.

A discriminant function for which the discriminant ratio is less than either of these values will be judged discordant with the data, at the corresponding level of significance.

The author is indebted to Prof. R. A. Fisher and to Prof. M. S. Bartlett for much helpful advice. The work described in this paper was carried out as part of the research programme of the Section of Mathematical Statistics, C.S.I.R.O.

REFERENCES

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.* **34**, 33.

Bartlett, M. S. (1939). A note on tests of significance in multivariate analysis. *Proc. Camb. Phil. Soc.* **35**, 180.

Bartlett, M. S. (1947*a*). The general canonical correlation distribution. *Ann. Math. Statist.* **18**, 1.

Bartlett, M. S. (1947*b*). Multivariate analysis. *J.R. Statist. Soc.* B, **9**, 176.

Bartlett, M. S. (1951). The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Ann. Eugen., Lond.*, **16**, 199.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc.* A, **139**, 343.

FISHER, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugen., Lond.*, **9**, 238.

FISHER, R. A. (1949). *The Design of Experiments*, 5th ed. Edinburgh: Oliver and Boyd.

HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika*, **28**, 321.

HSU, P. L. (1939). On the distribution of roots of certain determinantal equations. *Ann. Eugen., Lond.*, **9**, 250.

HSU, P. L. (1940). On the limiting distribution of the canonical correlations. *Biometrika*, **32**, 38.

NANDA, N. (1948). Limiting distribution of a root of a determinantal equation. *Ann. Math. Statist.* **19**, 340.

PITMAN, E. J. G. (1939 *a*). The estimation of location and scale parameters of a continuous population of any given form. *Biometrika*, **30**, 391.

PITMAN, E. J. G. (1939 *b*). Tests of hypotheses concerning location and scale parameters. *Biometrika*, **31**, 200.

ROY, S. N. (1939). p-Statistics or some generalizations in analysis of variance appropriate to multivariate problems. *Sankhyā*, **4**, 381.

ROY, S. N. (1942 *a*). The sampling distribution of p-statistics and certain allied statistics on the non-null hypothesis. *Sankhyā*, **6**, 15.

ROY, S. N. (1942 *b*). Analysis of variance for multivariate normal populations. *Sankhyā*, **6**, 35.

ROY, S. N. (1946). Multivariate analysis of variance: the sampling distribution of the numerically largest of the p-statistics on the non-null hypothesis. *Sankhyā*, **8**, 15.

WILKS, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471.

WILLIAMS, E. J. (1952). The interpretation of interactions in factorial experiments. *Biometrika* **39**, 65.

THE CONSTRUCTION OF BALANCED DESIGNS FOR EXPERIMENTS INVOLVING SEQUENCES OF TREATMENTS

By H. D. PATTERSON

Rothamsted Experimental Station

1. Introduction

This paper is concerned with the type of experimental design in which treatments are applied to the experimental material in a number of successive periods, each experimental unit receiving a different treatment in each period. Although some of the simpler examples of these designs have been in practical use in at least one field of research for several years, a systematic discussion of their general properties and the enumeration of possible designs does not appear to have been previously attempted.

In order that estimates of the effects of treatments and the errors can be estimated by a reasonably simple statistical analysis certain elements of balance, set out below, are required. The requirements depend largely upon which residual effects are to be estimated. For example, any Latin square arrangement, in which the rows represent the different periods so that a column of symbols refers to a sequence of treatments, is suitable for the case in which residual effects are negligible.

When first residual effects (i.e. the effects of treatments in the period after application) are to be estimated the class of available designs is more restricted. A simple example has been described by Cochran, Autrey & Cannon (1941) in connexion with feeding experiments on dairy cows. This arrangement, shown in Fig. 1, consists of two particular Latin squares.

Periods	Sequences					
I	1	2	3	1	2	3
II	2	3	1	3	1	2
III	3	1	2	2	3	1

Fig. 1. Change-over design for three treatments.

Arrangements of this sort are of limited value, as v, the number of treatments, is restricted to be equal to k, the number of periods. Designs for $v > k$ contain incomplete sequences and are related to the balanced incomplete block designs introduced by Yates (1936). For example, any Youden square can be used if residual effects are negligible.

Even if there are residual effects it is sometimes possible to use the less restricted designs, introducing intermediate periods with a uniform non-experimental treatment between the experimental periods. Estimates of residual effects are obtained from the measurements in the intermediate periods. Often, however, this type of arrangement cannot be used. It will not be considered further in this paper.

The application of change-over designs to animal feeding experiments has been surveyed by the present author (Patterson, 1951), but without particular reference to problems of construction. Reference may be made to this paper on questions of analysis.

2. General plan of the paper

In §§3 and 4 the conditions of balance are set out for the case in which first residual effects are estimated. It is shown how these conditions are related to the equations of estimation. A necessary but not sufficient relationship between numbers of treatments, periods and experimental units provides the basis of a table of possible designs for not more than six periods or sixty experimental units.

The remaining sections are devoted to the methods of construction. Many designs exist for which it is sufficient to define a number of leading sequences, the remaining sequences being obtained by a simple process. The construction of these designs is first discussed in general terms. Application is then made to:

(1) designs for any v equal to k, the number of periods, requiring v or $2v$ units depending on whether v is even or odd (§6);

(2) designs for v equal to a prime or power of a prime, any k not greater than v, and requiring $v(v-1)$ units (§7);

(3) designs for v a prime of the form $4n+3$, k equal to 3, and requiring $\frac{1}{2}[v(v-1)]$ units (§8);

(4) designs for v a prime of the form $4n+3$, k equal to $\frac{1}{2}(v+1)$ and requiring $2v$ units (§11);

(5) designs for v a prime of the form $4n+3$, odd k, and requiring $\frac{1}{2}[v(v-1)]$ units (§13).

A number of other designs are also obtained, and two of the listed three period designs are found to be non-existent.

Reference will be made several times in the course of this paper to various mathematical systems, and in particular to finite groups and finite fields. For convenience these systems, which contain finite numbers of elements, are briefly discussed here. In the case of groups some rule of combination, such as addition or multiplication, is defined. The combination of any two elements of the group, having regard to order, uniquely defines an element of the group. The group contains an identity which, when combined with any element of the group, leaves that element unaltered. To every element there corresponds an inverse, also in the group, such that the combination of the element and its inverse yields the identity. Finally, the associative law holds with respect to the rule of combination. If, in addition, the rule is commutative, the group is Abelian. The identities of additive and multiplicative groups can conveniently be represented by 0 and 1 respectively, and the inverses of an element a by $-a$ and a^{-1}. A multiplicative Abelian group which, together with an element 0 forms an additive Abelian group such that the distributive laws are satisfied, is called a finite field. It is known that finite fields must contain P^N elements, where P is any prime and N any positive integer.

3. Balanced designs for cases when residual effects are negligible

Whilst we are not primarily concerned with experiments in which residual effects are negligible, it is convenient to consider these first. A large number of arrangements is already available. These include Latin squares and Youden squares. The latter are balanced incomplete block arrangements, with the additional restriction that each treatment occurs just once in each position of the blocks. Extensions are also available in which each treatment occurs two, three, ... times in each position.

The conditions for balance can be stated as follows:

I. No treatment symbol occurs in a given sequence more than once.

II. Each symbol occurs in a given period an equal number of times.

III. Every two treatment symbols occur together in the same number of sequences. This number, in the usual notation, is λ.

A design consists of b sequences, and it is obvious that b/v must be integral. We have the restriction on b/v, implied by condition III, that

$$\frac{bk(k-1)}{v} \equiv 0 \pmod{v-1}. \tag{1}$$

A design does not necessarily exist for the minimum integral value of b/v satisfying (1).

Many of the problems involved in the construction of balanced incomplete block designs and the Youden-square type of design have been solved. A list of combinatorial solutions has been provided by Fisher & Yates (1948), and Bose (1939) has discussed a number of methods of construction.

4. Balanced designs when there are residual effects of treatments

The effects due to a treatment in periods after that of application are described as residual effects. In this paper we primarily consider the case in which:

(1) First residual effects only need be considered, i.e. the effect of the treatment on the measurement in the period two after the period of application can be neglected.

(2) Residual effects are independent of the treatment applied in the period in which they are observed. In this case direct and residual effects are said to be additive. For example, in the sequences 1, 2, 3; 2, 1, 3, the measurement of the first residual effect of treatment 1 is made in the second period of the first sequence and the third period of the second sequence. The treatments applied in these periods are treatments 2 and 3, but the residual effect is the same in both cases. If d is the direct effect of a treatment, and r the first residual effect, then the measurements in the two sequences are on d_1, d_2+r_1, d_3+r_2; d_2, d_1+r_2, d_3+r_1.

Denote by

$\bar{x}_{d1}$ the mean for treatment 1 in periods of application,

$\bar{s}_{d1}$ the mean for sequences in which treatment 1 is applied,

$\bar{x}_{r1}$ the mean in periods one after the application of treatment 1,

$\bar{s}_{r1}$ the mean for sequences in which treatment 1 is applied in one of the first $k-1$ periods where k is the number of periods.

If a design satisfies the conditions I–III and in addition conditions IV–VII set out below, estimates of differences are obtained in the form:

$$A(d_1-d_2)+B(r_1-r_2) = (\bar{x}_{d1}-\bar{x}_{d2})-(\bar{s}_{d1}-\bar{s}_{d2}), \tag{2}$$

$$B(d_1-d_2)+C(r_1-r_2) = (\bar{x}_{r1}-\bar{x}_{r2})-(\bar{s}_{r1}-\bar{s}_{r2}), \tag{3}$$

where A, B, C depend on the design.

Conditions IV–VII are:

IV. Each ordered succession of two treatment symbols should occur equally often in sequences. For example in Fig. 1 treatment 1 immediately follows each of treatments 2 and 3 at some stage.

V. Every two treatment symbols occur together in the same number of curtailed sequences formed by omitting the final period.

VI. In those sequences in which a given treatment occurs in the final period the other treatments occur equally often.

VII. In those sequences in which a given treatment occurs in any but the final period each other treatment occurs equally often in the final period.

Condition V is automatically satisfied with conditions III, VI and VII, but it should be noted that conditions I–V are not sufficient.

The following argument supports the statement that these conditions lead to equations of the type (2) and (3). Restrict the direct and residual effects so that

$$d_1+d_2+\dots d_v = 0,$$
$$r_1+r_2+\dots r_v = 0.$$

The conditions III–VII ensure that the only direct and residual effects involved in $\bar{x}_{d1}, \bar{x}_{r1}, \bar{s}_{d1}$ and $\bar{s}_{r1}$ are d_1 and r_1. The following diagram shows which conditions are effective in this respect:

	Effects involved	
	d_1	r_1
$\bar{x}_{d1}$	.	IV
$\bar{x}_{r1}$	IV	.
$\bar{s}_{d1}$	III	III, VII
$\bar{s}_{r1}$	V, VII	V

Thus condition IV ensures that $\bar{x}_{d1}$ is free of all residual effects but r_1, and $\bar{x}_{r1}$ is free of all direct effects but d_1.

The restriction on b/v follows from condition IV and is that

$$\frac{b(k-1)}{v} \equiv 0 \,(\mathrm{mod}\, v-1). \tag{4}$$

Designs have been sought for $k = 3, 4, 5, 6$ requiring 60 or fewer units. Minimum values of b/v obtained from (4) are given in Table 1, but it does not necessarily follow that designs are available for this minimum value.

Table 1. *Designs balanced for first residual effects*

v	$k=3$		$k=4$		$k=5$		$k=6$	
	min. b/v	Section	min. b/v	Section	min. b/v	Section	min. b/v	Section
3	2	6, 7	—	—	—	—	—	—
4	—	—	1	6	—	—	—	—
5	2	† 9	4	7	2	6	—	—
6	5	*	—	—	5	*	1	6
7	3	5, 8	2	11	3	13	6	7
8	7	7	7	7	7	7	7	7
9	4	† 9	—	—	2	*	—	—
10	—	—	3	*	—	—	—	—
11	5	8	—	—	5	13	2	11
13	—	—	4	10	3	12	—	—
16	—	—	—	—	—	—	3	*

k, no. of periods. v, no. of treatments. b, no. of units.

* No design found. † Design non-existent.

Designs are also available for

v	k	b/v	Section
6	3	10	10
10	4	6	10

In addition to the designs of Table 1, a number outside the range have been obtained and are mentioned in the appropriate sections.

5. Method of construction using differences

Consider a design for seven treatments and three periods such as that based on three Latin rectangles and shown in Fig. 2. The treatments are represented by the non-negative residues (mod 7). Reference to the conditions I–VII shows that the design is balanced. The structure is of considerable interest, each sequence of a given rectangle being obtained from the previous sequence by replacing symbols in the order 0 1 2 3 4 5 6 0. A solution of this sort is described as a cyclic solution. It should be noted that for any pair of rows in a given rectangle, the difference between treatment symbols in every sequence is the same.

Periods	Sequences		
I	0 1 2 3 4 5 6	0 1 2 3 4 5 6	0 1 2 3 4 5 6
II	5 6 0 1 2 3 4	3 4 5 6 0 1 2	6 0 1 2 3 4 5
III	6 0 1 2 3 4 5	5 6 0 1 2 3 4	3 4 5 6 0 1 2
	1st rectangle	2nd rectangle	3rd rectangle

Fig. 2. Three period change-over design for seven treatments, based on Latin rectangles.

In this section the general problem of obtaining solutions for which it is sufficient to list b/v initial sequences, or sets of differences, is discussed. Consider a design consisting of b/v Latin rectangles each having k rows and v columns. Represent each treatment by one of the elements of an additive Abelian group of order v. One treatment may therefore be represented by 0. The rectangles are of special type such that the differences between successive rows of the ith rectangle are $\delta_{1i}, \delta_{2i}, \ldots, \delta_{k-1,i}$, $i = 1, 2, \ldots, b/v$. The sets of differences, one set to each rectangle, must be such that

(1) No two treatments of the leading sequence in the ith rectangle

$$0,\ \delta_{1i},\ \delta_{1i}+\delta_{2i},\ \ldots,\ \delta_{1i}+\delta_{2i}+\ldots+\delta_{k-1,i},\quad i = 1, 2, \ldots, b/v$$

are the same. With the method of construction of the remaining sequences given below this condition ensures that Latin rectangles are obtained.

(2) Each difference is some non-zero element of the additive Abelian group of treatment symbols. The δ's do not necessarily define the group apart from the identity; thus some of the δ's may be the same. As yet we do not even say that all the non-zero elements of the group are represented in the δ's, although it will be shown later that this is a necessary condition for balance. The remaining sequences of a rectangle are obtained by adding the non-zero elements of the group of treatment symbols in turn to the symbols of the leading sequence, keeping the same order.

It will be convenient in the following discussion to consider the sequences which include 0. These sequences are given by the k columns of the array shown below and are obtained by adding the k elements $0, -\delta_{1i}, -(\delta_{1i}+\delta_{2i}), \ldots, -(\delta_{1i}+\delta_{2i}+\ldots+\delta_{k-1,i})$ in turn to the elements of the leading sequence. An example of such an array is given by the first, third and fifth columns of the second rectangle of Fig. 2.

$$\begin{array}{ccccc}
0 & -\delta_{1i} & -(\delta_{1i}+\delta_{2i}) & \ldots & -(\delta_{1i}+\delta_{2i}+\ldots+\delta_{k-1,i}) \\
\delta_{1i} & 0 & -\delta_{2i} & \ldots & -(\delta_{2i}+\ldots+\delta_{k-1,i}) \\
\delta_{1i}+\delta_{2i} & \delta_{2i} & 0 & \ldots & -(\delta_{3i}+\ldots+\delta_{k-1,i}) \\
\delta_{1i}+\delta_{2i}+\delta_{3i} & \delta_{2i}+\delta_{3i} & \delta_{3i} & \ldots & -(\delta_{4i}+\ldots+\delta_{k-1,i}) \\
\vdots & \vdots & \vdots & & \vdots \\
\delta_{1i}+\delta_{2i}+\ldots+\delta_{k-1,i} & \delta_{2i}+\delta_{3i}+\ldots+\delta_{k-1,i} & \delta_{3i}+\ldots+\delta_{k-1,i} & \ldots & 0
\end{array}$$

A number of conditions must be imposed on the choice of the δ so that the conditions for balance are satisfied. Conditions I and II are obviously satisfied as Latin rectangles are used. Denote by δ_1, for example, all the elements $\delta_{1i}, i = 1, \ldots, b/v$, with the convention that any operation on the δ_1 is to be interpreted as carried out on each δ_{1i}. Thus the set $\delta_1+\delta_2$ is to be interpreted as the b/v elements $\delta_{1i}+\delta_{2i}$ taken over all i, and not, as is usual in the notation of sets, the set of elements belonging to at least one of the sets δ_1 and δ_2. Similarly, denote the differences $\delta_{1i}, \delta_{2i}, \ldots, \delta_{k-1,i}$ by α_i, and the set of all the α_i by α. Also let the set of all the elements to the left of the leading diagonal of the above diagram be denoted by β_i, and the set of all the β_i by β. The set of cumulative sums starting with $\delta_{k-1,i}$, i.e.

$$(\delta_{k-1,i}, \delta_{k-2,i}+\delta_{k-1,i}, \ldots, \delta_{1i}+\delta_{2i}+\ldots+\delta_{k-1,i}),$$

is denoted by γ_i, and the set of γ_i by γ. Condition III is satisfied only if $+\beta$ and $-\beta$ together include all the non-zero elements of the group an equal number of times. In this case all pairs of treatment symbols, one of which is 0, occur in an equal number, λ, of sequences. Condition IV is satisfied only if all the treatments immediately following 0 in sequences, i.e. the elements in α, include each non-zero element of the group an equal number of times. Similarly, all the sequences in which 0 occurs in the final period contain as the other entries the elements in $-\gamma$ and the sequences in which 0 occurs in any but the final period contain as entries in the final period the elements in γ. It follows that Conditions VI and VII are satisfied only if the elements in γ include each non-zero element of the group an equal number of times. All the new conditions are sufficient, as sequences involving an element other than 0 are obtained by adding this element to each of the entries including 0 in the above diagram of sequences. Thus the treatments immediately following w, where w is an element of the group, are given by adding w to each element in α, and can be represented by $\alpha+w$. If α includes each element of the group equally often then so must $\alpha+w$.

It is convenient in testing the balance of a proposed arrangement to write down b/v arrays of symbols, one for each Latin rectangle:

$$\begin{array}{cccc}
\delta_{1i} & \delta_{1i}+\delta_{2i} & & \\
\delta_{2i} & \delta_{2i}+\delta_{3i} & \delta_{1i}+\delta_{2i}+\delta_{3i} & \\
\delta_{3i} & \vdots & \vdots & \ldots \quad \delta_{1i}+\delta_{2i}+\ldots+\delta_{k-1,i} \\
\vdots & \delta_{k-2,i}+\delta_{k-1,i} & & \\
\delta_{k-1,i} & & &
\end{array}$$

Such an array is described here as a triangular diagram. It is then possible to identify α, β, γ without much difficulty.

Examples of finite additive Abelian groups are provided by the systems of non-negative residues (mod v), where v is any positive integer. For the design of Fig. 2 in which $v = 7$ the triangular diagrams are:

5			3			6		
	6			5			3	
1			2			4		

1st rectangle 2nd rectangle 3rd rectangle

α, γ are easily identified and each seen to include all the non-zero residues just once. The elements in β are (5, 1, 6, 3, 2, 5, 6, 4, 3) and the elements in $-\beta$ are therefore (2, 6, 1, 4, 5, 2, 1, 3, 4), so that $+\beta$ and $-\beta$ together include each non-zero residue 3 times.

Whilst, for the sake of simplicity, finite additive Abelian groups have been used in the above discussion it is possible to consider other mathematical systems (not necessarily satisfying all the postulates of groups). To take an obvious example a multiplicative group could be used, in which case multiplication of the treatment symbols in the leading sequences by elements of the group other than the identity would yield the remaining sequences.

6. Designs based on complete Latin squares

Some designs can be obtained from complete Latin squares. The number of treatments is then restricted to be the same as the number of periods. Represent the treatments by the non-negative residues (mod v), where v is any positive integer. Williams (1949) has shown that for even $v = k$ when the minimum value of b/v is 1 the differences

$$1, v-2, 3, v-4, \ldots, 2, v-1$$

provide a cyclic solution. For odd $v = k$ when the minimum value of b/v is 2, the two squares are defined by the differences

$$1, v-2, 3, v-4, 5, \ldots, v-2, 1,$$
$$v-1, 2, v-3, 4, v-5, \ldots, 2, v-1.$$

These are fairly straightforward cases. Inspection shows that the differences, α, for the above designs include each of the non-zero residues (mod v) equally often. Latin squares and Latin squares with one missing row are known to be balanced in the sense demanded by conditions III, V and VI. In any case the contents of β and $-\beta$ together and γ can be checked as in the example of the previous section.

7. Designs based on complete sets of orthogonal Latin squares

In this section it is shown that designs can be constructed for any v which is a prime or a power of a prime, any $k \leqslant v$ and a value of b/v, which is not necessarily the minimum value, given by $v-1$. An example of such a design is given in Fig. 3.

Periods	Sequences		
I	1 2 3 4	1 2 3 4	1 2 3 4
II	2 1 4 3	3 4 1 2	4 3 2 1
III	3 4 1 2	4 3 2 1	2 1 4 3

Fig. 3. Change-over design for four treatments and three periods.

This design which, on inspection, is seen to be balanced was obtained from the first three rows of the three orthogonal 4×4 Latin squares listed by Fisher & Yates (1948).

Consider a set of $v-1$ orthogonal $v \times v$ Latin squares obtained by a field method and such that any square of the set is obtained from any other square of the same set by a permutation of rows. Such a set exists for all values of v which are primes or powers of primes. The balanced designs are obtained by taking k corresponding rows from each of the $v-1$ squares, the columns of symbols then defining the treatment sequences.

The proof depends on the construction of the orthogonal squares as described by Stevens (1939). Consider a field of v symbols: $0, 1, u_2, \ldots, u_{v-1}$, where $v = P^N$, P is a prime and N any positive integer. A Latin square of symbols is written as

$$u_\lambda\{u_x + u_y\} \quad (u_\lambda \neq 0), \tag{5}$$

where u_x defines the row, u_y the column and u_x, u_y take the values $0, 1, u_2, \ldots, u_{v-1}$. For a given square u_λ is constant, but if it is allowed to take the values $1, u_2, \ldots, u_{v-1}$ in different squares a set of orthogonal squares is defined. Consider any pair of rows $u_x, u_{x'}$ $(u_x \neq u_{x'})$. The differences between the symbols of these rows in the $v-1$ squares are

$$u_\lambda(u_x - u_{x'}),\ u_\lambda = 1,\ u_2,\ \ldots,\ u_{v-1}, \tag{6}$$

and these, by definition of the field, must include each non-zero element of the field just once. It follows that if we take any number of corresponding rows the symbols in a given position of the triangular diagrams taken over all i (i.e. for all the squares) include each non-zero element of the field just once. Thus, for example, there is a one-to-one correspondence between the elements in $(\delta_{11}+\delta_{21}, \delta_{12}+\delta_{22}, \delta_{13}+\delta_{23}, \ldots, \delta_{1,v-1}+\delta_{2,v-1})$ and the elements in $(1, u_2, \ldots, u_{v-1})$. The conditions I–VII for a balanced design are therefore satisfied. It can be seen that this design is also balanced for any residual effects (second, third, ..., $k-1$th) which can be estimated.

It is of some interest to note that more general abstract algebras can be used. The algebra of finite fields is associative and commutative with respect to addition and multiplication and satisfies both distributive laws, e.g.

$$u_\lambda(u_x - u_{x'}) = u_\lambda u_x - u_\lambda u_{x'},$$
$$(u_x - u_{x'})\,u_\lambda = u_x u_\lambda - u_{x'} u_\lambda.$$

It appears that of the laws just mentioned it is sufficient to postulate one of the distributive laws. Orthogonal sets of Latin squares and change-over designs can then be obtained as before from (5).

It is important to recognize that the above proof does not extend to all sets of orthogonal Latin squares. For example, Bose & Nair (1941) have given a set of squares associated with a non-Desarguesian geometry. This set does not have the property by which each square may be obtained from any other by a permutation of rows. Not all the arrangements obtained by taking k corresponding rows from each of the $v-1$ squares of the set are balanced.

In the case of a design based on all the $v!/(v-k)!$ possible sequences represent the treatments by the group of least positive residues (mod v). The differences between any two rows must include each of the non-zero residues equally often so that this design is also balanced for up to $k-1$th residual effects. These arrangements have, however, the disadvantage that large numbers of experimental units are required.

8. Three-period designs

The results of previous sections will now be used in finding a number of three-period designs. A general series is available for any number of treatments which is a prime or power of a prime of the form $4n+3$, where n is a positive integer, with b/v taking the minimum value of $\frac{1}{2}(v-1)$. A number of designs for other values of v can also be obtained.

Consider a design with differences in the ith rectangle given by $(\delta_{1i}, \delta_{2i})$ $(i = 1, 2, \ldots, b/v)$. Let the treatments be elements of the group of non-negative residues (mod v). As α, the set consisting of δ_1 and δ_2 and γ, the set of δ_2 and $\delta_1+\delta_2$ must each include all non-zero residues equally often, there is a one-to-one correspondence between the elements in δ_1 and the elements in $\delta_1+\delta_2$. Further, for any odd v the sum of all the elements in $\delta_1+\delta_2$ must be zero, so that the sums of the elements in δ_1 and δ_2 must each be zero. For odd v and a minimum value of b/v, i.e. $\frac{1}{2}(v-1)$, the δ_1, δ_2 should be chosen so that there is a one-to-one correspondence between $-\delta_1$ and δ_2. The following sets of elements illustrate the above remarks in the case $v = 7$, $b/v = 3$.

δ_1	5	3	6
δ_2	1	2	4
$\delta_1+\delta_2$	6	5	3

This design was given in Fig. 2.

Designs have been obtained in this way for $v = 7, 11, 13, 19, 21$, and there is more than one solution for each design. Those for $v = 13$ are given by the pairs of differences (with alternatives for the elements of δ_2):

(1)	δ_1	1	2	3	9	5	6	(2)	δ_1	12	2	10	4	5	6
	δ_2	8	4	11	7	10	12		δ_2	11	3	7	8	1	9
	or	4	7	12	10	11	8		or	3	8	9	1	7	11
(3)	δ_1	1	11	3	9	8	7	(4)	δ_1	4	7	8	10	11	12
	δ_2	10	5	4	12	6	2		δ_2	6	1	3	2	9	5
	or	2	10	6	5	12	4		or	3	5	2	1	6	9

A class of designs can be obtained for certain odd v equal to a prime or power of a prime by a method involving the use of the properties of finite fields. This is the general series previously mentioned. Each treatment is represented by an element of a field of v elements. The elements of the field can be represented by

$$0, 1, x, x^2, \ldots, x^{v-2}, \tag{7}$$

where x is a primitive root of the equation $x^{v-1} = 1$, i.e. x is such that none of $x, x^2, \ldots, x^{v-2}$ is equal to 1. We have to divide the non-zero elements into $\frac{1}{2}(v-1)$ pairs defining the differences between rows in each rectangle. Consider the pairs

$$\left.\begin{array}{llllll} \delta_1 & x^0 & x^2 & x^4 & \ldots & x^{v-3}, \\ \delta_2 & x^1 & x^3 & x^5 & \ldots & x^{v-2}. \end{array}\right\} \tag{8}$$

Each non-zero element is represented once so that condition IV is satisfied. γ is represented by $x^1, x^3, x^5, \ldots, x^0(x+1), x^2(x+1), \ldots$. Now $x+1$ must be either an even or odd power of x. If it is an even power γ includes all the non-zero elements of the field. If it is an odd power we must use the pairs in the reverse order

$$\left.\begin{array}{llllll} \delta_1 & x^1 & x^3 & x^5 & \ldots & x^{v-2}, \\ \delta_2 & x^0 & x^2 & x^4 & \ldots & x^{v-3}, \end{array}\right\} \tag{9}$$

when γ is $x^0, x^2, x^4, \ldots, x^0(x+1), x^2(x+1), \ldots$, and includes all the non-zero elements. In any case conditions VI and VII can be satisfied by taking one or other of (8) and (9).

In order that conditions III and hence V may be satisfied $+\beta$ and $-\beta$ together must include all the non-zero elements three times. If $+\gamma$ and $-\gamma$ together include these twice the elements

$$x^0, x^2, \ldots, -x^0, -x^2, \ldots$$

must include each non-zero element just once. As

$$x^{\frac{1}{2}(v-1)} = -1,$$

a property which follows from the fact that

$$(x^{\frac{1}{2}(v-1)} - 1)(x^{\frac{1}{2}(v-1)} + 1) \equiv 0,$$

this will, with one exception, obviously be the case if $\frac{1}{2}(v-1)$ is odd. The exception arises in the case $v = 3$ when $x^0(1+x)$ is $\mathfrak{C}$. The above method gives a design for any odd v which is a prime or a power of a prime, $k = 3$ and odd $b/v = \frac{1}{2}(v-1)$, i.e. v is of the form $4n+3$ where n is a positive integer. In general, solutions are found to be available from the differences in some order:

$$\left.\begin{array}{llllll} \delta_1 \text{ or } \delta_2 & x^0 & x^2 & x^4 & \ldots, \\ \delta_2 \text{ or } \delta_1 & x^i & x^{i+2} & x^{i+4} & \ldots, \end{array}\right\} \tag{10}$$

where i is odd and not equal to $\frac{1}{2}(v-1)$, and the powers are reduced where necessary, using the relationship

$$x^{v-1} = 1.$$

For $v = P^N$ and $N = 1$ we obtain a cyclic solution, for $N = 2$ a dicyclic solution and in general an N-cyclic solution.

The following cases are considered here:

(1) $v = 7$. A primitive element is 3 so that the power cycle $3^0, 3^1, 3^2, \ldots$ (reduced mod 7) consists of the integers $1, 3, 2, 6, 4, 5$. $x+1$ is 4, i.e. 3^4, and as $\frac{1}{2}(v-1)$ is odd conditions III and V are satisfied. From the pairs (8) we find the design given by the differences

$$\begin{array}{lllll} \delta_1 & 1 & 2 & 4 \\ \delta_2 & 3 & 6 & 5 \end{array}$$

which is the alternative solution to that given previously.

(2) $v = 11$. A primitive element is 2, $x+1$ is 3 and therefore equal to x^8, and $\frac{1}{2}(v-1)$ is odd. Hence from (8) the differences are

$$\begin{array}{lllllll} \delta_1 & 1 & 4 & 5 & 9 & 3 \\ \delta_2 & 2 & 8 & 10 & 7 & 6 \end{array}$$

(3) $v = 13$. Here $\frac{1}{2}(v-1)$ is even and the method fails, but solutions have been given earlier.

(4) $v = 3^3$. This design, of little practical value, is of some theoretical interest, as we have not so far considered an example of a power of a prime. The field can be represented by the elements (7) together with the relationship $x^3 = x+2$, and the reduction of coefficients mod 3. Thus x^4 is $x(x+2)$, i.e. x^2+2x, and may conveniently be denoted by 120. We find $x+1 = x^9$ so must use (9) rather than (8). The solution is tricyclic and may be represented:

$$\begin{array}{lllllllllllll} \delta_1 & 010, & 012, & 212, & 122, & 011, & 112, & 002, & 200, & 210, & 222, & 101, & 220, & 201 \\ \delta_2 & 001, & 100, & 120, & 111, & 202, & 110, & 102, & 020, & 021, & 121, & 211, & 022, & 221 \end{array}$$

These pairs of differences give the leading sequences, the remainder being obtained by adding each non-zero element of the field, here represented by the non-zero residues (mod, 3, 3, 3) in turn.

9. Non-existence of two three-period designs

There are still a number of important gaps in the table of three-period designs not covered by the previous section. It has been found that designs do not exist for $v = 5$, $v = 9$ with values of b/v which are minima.

Consider any three-period design with odd v and a minimum value of b/v. In the first place there are just $\frac{1}{2}v(v-1)$ sequences, so that the only possible way in which condition V can be satisfied is to have each possible pair of symbols just once in the curtailed sequences of the first two periods. Thus if 12. is one sequence then 21. cannot be another. (The symbol . here refers to any treatment, and not necessarily to the same treatment each time it is used.) It follows from condition IV that .21 is included and .12 excluded. In this case a further sequence is 1.2 as 2.1 is not allowable (condition VI).

Now consider the design $v = 5$, $b/v = 2$. Number the treatments from 1 to 5. Without any loss of generality we may write, referring to condition VI, one pair of sequences as

1 3
2 4
5 5

These are the only sequences in which 5 may occur in the final period, and the remaining treatments must include each other symbol just once. Now 5 has immediately followed both 2 and 4, so that another pair of sequences must be

1 3
5 5
. .

Further, 1 must come immediately before each of 2, 5, 3, 4 so that the sequences

(*v*) (*w*)
1 1
3 4

are required where (*v*), (*w*) are 2 and 5 in some order. But if (*v*) is either 2 or 5 condition V cannot be satisfied as the sequences 12. and 15. have already occurred. It follows that the design is non-existent. The design for the case $v = 9$, $k = 3$, $b/v = 4$ is also non-existent. This can be demonstrated in much the same way, but the proof is lengthy and is omitted.

10. A second method of construction

No design has been found for $v = 6, k = 3, b/v = 5$, but one can be obtained for $b/v = 10$, by a method described by the author elsewhere (Patterson, 1951). This design is based on the balanced incomplete block design for six varieties in ten blocks of three. Two orthogonal 3×3 Latin squares, themselves forming a balanced design, are constructed for the treatment symbols of each block. Thus two of the incomplete blocks might be written (1, 2, 3) and

(2, 4, 5), in which case twelve of the sequences of the final design would be given by the columns of

1	2	3	1	2	3	2	4	5	2	4	5
2	3	1	3	1	2	4	5	2	5	2	4
3	1	2	2	3	1	5	2	4	4	5	2

Such designs, which usually, however, do not have minimum values of b/v, can be constructed where any balanced incomplete block design exists, superimposing on each block either one or two Latin squares of the type described in §6, depending on whether k is even or odd.

This method may be used in the two cases $v = 7$ and $v = 13$ with $k = 4$, as 4×7 and 4×13 Youden squares exist. For these designs $b/v = 4$ and this is a minimum value in the second case. When $v = 13$ the solution may be represented in terms of four sets of three differences in the additive group of non-negative residues (mod 13). An example is

δ_1	1	8	10	7
δ_2	2	4	9	11
δ_3	6	3	5	12

As a balanced incomplete block design exists for 10 varieties in 15 blocks of 4 a design balanced for first residual effects can be obtained for $b/v = 6$. The design cannot be arranged as 6 separate 4×10 rectangles but can be divided into 4×4 squares. This point will be raised later.

11. A SERIES OF DESIGNS FOR $k = \frac{1}{2}(v+1)$, $b/v = 2$

In the previous section a design was given for $v = 7$, $k = 4$, $b/v = 4$. The minimum value of b/v is, however, $\min b/v = 2$. Consider the set of differences 1, 2, 3 and the set given by the inverses, i.e. 6, 5, 4, in the additive group of non-negative residues (mod 7). All the conditions for balance are satisfied, each non-zero residue being included in α just once, in β and $-\beta$ together just four times, and in γ just once. The design is given in Fig. 4.

Periods	Sequences													
I	0	1	2	3	4	5	6		0	1	2	3	4	5	6
II	1	2	3	4	5	6	0		6	0	1	2	3	4	5
III	3	4	5	6	0	1	2		4	5	6	0	1	2	3
IV	6	0	1	2	3	4	5		1	2	3	4	5	6	0

Fig. 4. Four period change-over design for seven treatments, requiring fourteen units.

It is of interest to note that similar properties hold for some other prime v. For any such prime there is more than one solution. In fact a balanced design with $k = \frac{1}{2}(v+1)$, $b/v = 2$ exists for any prime v equal to $4n+3$ where n is a positive integer. Treatments are represented by the elements of the group of non-negative residues (mod v). The sets of differences are

$$\{2s, 4s, 6s, \ldots, (v-1)s\} \quad \text{and} \quad \{-2s, -4s, -6s, \ldots, -(v-1)s\}, \text{mod}\,(v), \qquad (11)$$

where s is a positive integer not a multiple of v.

It is obvious that α includes each non-zero residue just once. The set β of all elements in the two triangular diagrams can be separated into β' and $-\beta'$ as the elements of one diagram consist of the inverses of the corresponding elements of the other. The elements in β' can be written as the triangular diagram:

$$\begin{array}{ccccc} & & 2s & & \\ & & & 6s & \\ & 4s & & & 12s \\ & & 10s & & \vdots \\ 6s & & & & 3(v-2)s \ \ldots \\ \vdots & & 2(v-2)s & & \\ (v-1)s & & & & \end{array}$$

There are simple relationships between the elements in β'. It can be shown that

$$(i+j-1)+(i+j-3)+\ldots+(i-j+1)\equiv ij \text{ for } i>j.$$

If, therefore, we take values of i, j such that $i+j$ is odd, $\leqslant v$ and $i>j$, each element of β' can be defined. We are led to consider the matrix A equal to (a_{ij}), where $a_{ij}=ijs \pmod v$ $(i,j=1,2,\ldots,v-1)$, and v is an odd prime. Thus

$$A=\begin{pmatrix} s & 2s & \ldots & (v-1)s \\ 2s & 4s & \ldots & 2(v-1)s \\ 3s & 6s & \ldots & 3(v-1)s \\ \vdots & \vdots & & \vdots \\ (v-1)s & 2(v-1) & \ldots & (v-1)^2 s \end{pmatrix} \pmod v.$$

Certain properties of this matrix are noted. As $a_{ij}\equiv a_{ji}$ it is obviously symmetrical about the principal diagonal. Further

$$ij\equiv i'j' \pmod v,$$

where

$$i'=v-i, \quad j'=v-j,$$

so that A is symmetrical about the secondary diagonal (i.e. $a_{ij}=a_{i'j'}$). As

$$\left.\begin{aligned} ij &\equiv -i'j \pmod v, \\ ij &\equiv -ij' \pmod v, \end{aligned}\right\} \tag{12}$$

we have

$$a_{ij}\equiv -a_{i'j}\equiv -a_{ij'} \pmod v.$$

Each of i and j include all the non-zero residues just once and therefore so does each row and column of A.

Now construct the matrix B equal to (b_{ij}), where

$$\begin{aligned} b_{ij} &= a_{ij} \quad (i+j \text{ odd}) \\ &= 0 \quad (i+j \text{ even}). \end{aligned}$$

Using (12) it follows that the elements of any row or column of B and their inverses include all the non-zero residues just once. The leading diagonal of B consists of $0, 0, 0, \ldots$, so that if we define the matrix C equal to (c_{ij}), where

$$\begin{aligned} c_{ij} &= b_{ij} \quad (i \geqslant j) \\ &= 0 \quad (i<j), \end{aligned}$$

we find that $+C$ and $-C$ together include each non-zero residue just $\frac{1}{2}(v-1)$ times.

Some properties of quadratic residues R, and quadratic non-residues N, are now required.

(1) A number R is a quadratic residue of a prime v if there exists some r such that

$$r^2 \equiv R \pmod{v}.$$

The least positive residues of the squares

$$1^2, 2^2, 3^2, \ldots, \left(\frac{v-1}{2}\right)^2$$

are different and yield all the quadratic residues.

(2) Euler's criteria for quadratic residues and non-residues are

$$R^{\frac{1}{2}(v-1)} \equiv 1 \pmod{v},$$
$$N^{\frac{1}{2}(v-1)} \equiv -1 \pmod{v}.$$

It follows that if v is of the form $4n+3$ then $-R$ is a quadratic non-residue.

(3) The product of two quadratic non-residues is a quadratic residue.

(4) The product of a quadratic residue and a quadratic non-residue is a quadratic non-residue.

The above results can be proved quite easily if it is realized that the quadratic residues consist of the even powers of the primitive elements.

For the remainder of the discussion we take v in the form $4n+3$. Consider the non-zero elements of the secondary diagonal of C. These can be shown as

$$\gamma' = \left(-1^2 s, -2^2 s, -3^2 s, \ldots, -\left\{\frac{v-1}{2}\right\}^2 s\right).$$

Using the above properties we find that γ' includes all the quadratic residues or quadratic non-residues of v according to whether s is a quadratic non-residue or quadratic residue. From property (2) $-\gamma'$ includes the remaining non-zero residues.

Finally, define the matrix D, equal to (d_{ij}), such that

$$d_{ij} = c_{ij} \quad (i+j \leqslant v)$$
$$= 0 \quad (i+j > v).$$

Thus

$$D = \begin{pmatrix} 0 & 0 & 0 & \ldots \\ 2s & 0 & 0 & \ldots \\ 0 & 6s & 0 & \\ 4s & 0 & 12s & \\ 0 & 10s & 0 & \\ \vdots & \vdots & \vdots & \\ \vdots & 2(v-2)s & \vdots & \\ (v-1)s & 0 & 0 & \ldots \end{pmatrix} \pmod{v}$$

The secondary diagonal is the same as that of C, and as C is symmetrical about this diagonal D and $-D$ together include each non-zero residue $n+1$ times. The non-zero elements in D are the elements in β'. It has been shown that α and similarly γ' and $-\gamma'$ together, i.e. γ, include each non-zero residue just once. The differences (11) therefore give a balanced design. Different solutions are given by taking $s = 1, 2, \ldots, \frac{1}{2}(v-1)$.

It is worth noting that the existence of such designs implies the existence of a $v \times \frac{1}{2}(v+1)$ and hence also a $v \times \frac{1}{2}(v-1)$ Youden square for prime v of the form $4n+3$, a result previously obtained by Bose (1939).

12. The design for $v = 13$, $k = 5$, $b/v = 3$

A design is available for 13 treatments and 5 periods requiring 39 units. Let the treatments be represented by the elements of the field of 13 elements. The differences in the three rectangles are

δ_1	x^0	x^4	x^8
δ_2	x^1	x^5	x^9
δ_3	x^2	x^6	x^{10}
δ_4	x^3	x^7	x^{11}

The primitive root is $x = 2$, so that $1+x = x^4$, $1+x+x^2 = x^{11}$, $1+x+x^2+x^3 = x$. Each non-zero element is included once in α, and using the above relationships and $x^{12} = 1$, $x^6 = -1$, we find that each non-zero element is included once in γ and five times in β and $-\beta$ together. The differences giving a balanced design are therefore:

δ_1	1	3	9
δ_2	2	6	5
δ_3	4	12	10
δ_4	8	11	7

13. Further designs for odd numbers of periods

It is not difficult to extend the series of three-period designs for values of v which are primes or powers of primes of the form $4n+3$ to other odd numbers of periods. As before represent the treatments by the elements of a field of v elements. The method of construction is to write down sets of differences satisfying certain conditions:

(1) The first set should contain even and odd powers of x, the primitive element, equally often.

(2) The cumulative sums in the first set, starting with the last difference, should contain even and odd powers of x equally often.

(3) The leading sequence obtainable from the differences in the first set contains no treatment more than once.

(4) The remaining sets of differences, making a total of $\frac{1}{2}(v-1)$, are obtained by multiplying the elements of the first set by each of the even powers of x in turn.

Such a method is similar to that used in constructing designs from sets of orthogonal Latin squares, the difference being that u_λ takes values from only half the non-zero elements of the field.

It can easily be shown that such sets of differences lead to a balanced design for an odd number of periods. The proof depends on the property that for the values of v under consideration the inverses (with respect to addition) of the even powers of x are odd powers of x. If there are $v-1$ differences in each set $\frac{1}{2}(v-1)$ mutually orthogonal Latin squares are obtained. In this case conditions VI and VII are automatically satisfied so that there is no need to consider the cumulative sums.

Four examples are given below. These are such that the symbols of each three successive periods themselves define a balanced three-period design. In each case the first set of differences is given, the remaining sets being obtained by multiplying the differences of the first set by the quadratic residues other than 1:

(1) $v = 7, k = 5, b/v = 3$ (5, 1, 3, 2)

quadratic residues of 7: 1, 2, 4

(2) $v = 11, k = 5, b/v = 5$ (1, 8, 5, 7)

(3) $v = 11, k = 7, b/v = 5$ (1, 2, 5, 7, 3, 6)

(4) $v = 11, k = 9, b/v = 5$ (1, 8, 5, 7, 3, 2, 4, 10)

quadratic residues of 11: 1, 3, 4, 5, 9.

14. General remarks

Referring to simple change-over trials Cochran *et al.* (1941) have shown that there is some advantage in using blocks of cows, each block corresponding to a single square. This procedure may often be useful with other types of experimental material. Most of the designs of this paper, including all those found by the method of differences, can be arranged in blocks of v units.

In the cases in which the balanced designs with $v = k = 3, b/v = 2$ or $v = k = 4, b/v = 1$ are superimposed on each block of a balanced incomplete block arrangement the size of blocks can always be equal to k. Thus the design $v = 13, k = b/v = 4$ can be arranged in blocks of 4 units. Some of these designs can also be arranged in blocks of v units, but there are exceptions. For example, the arrangement for $v = 10$, $k = 4$, $b/v = 6$ is based on a non-resolvable incomplete block design and cannot be arranged in six rectangles.

A word of warning is necessary on the analysis connected with the designs of this paper. The straightforward analysis following from the least squares method of fitting constants including constants for residual effects is not appropriate under all systems of correlations between experimental error terms. In practice, with certain types of material, it seems likely that no serious error is made if this point is ignored. A discussion of the analysis and of advantages and disadvantages obtained in using change-over trials is given by Patterson (1950, 1951).

There is also a large number of partially balanced designs. The analysis is rather more complicated. It is not proposed to discuss these designs here.

15. Designs balanced for first and second residual effects

A design is balanced for first and second residual effects if, in addition to conditions I–VII, the arrangement satisfies

VIII. Each treatment is preceded, two periods earlier, by each other treatment equally often.

IX. Each ordered pair of treatments occurs at the end of an equal number of sequences.

X. In those sequences in which a given treatment occurs in the last but one period the other treatments occur equally often.

XI. In those sequences in which a given treatment occurs in any but the last but one period the other treatments occur equally often in the last but one period.

The restrictions on b/v are, from conditions IV and VIII,

$$b(k-1)/v \equiv 0 \pmod{v-1},$$
$$b(k-2)/v \equiv 0 \pmod{v-1},$$

so that by subtracting $$b/v \equiv 0 \pmod{v-1}.$$

It follows that b/v cannot be less than $v-1$. Condition IX, which is less obvious, implies the same thing.

If a design with $v = k$ satisfies conditions I and II, then conditions V–VII, X, XI are automatically satisfied. Williams (1950) has shown that the construction of balanced designs in the cases $v = k$ is related to the problem of finding $\frac{1}{2}[v(v-1)]$ arrangements of $v+1$ objects in a circle such that no object has the same pair of neighbours in any two circles. Solutions to this problem exist for general values of v.

Designs balanced for pairs of residual effects will not be considered further in this paper. It should be noted, as remarked in §7, that solutions are provided for v prime or a power of a prime and any k such that $v \geqslant k \geqslant 3$ by designs involving rectangles from complete sets of orthogonal Latin squares.

Summary

This paper deals with the construction of designs for experiments in which sequences of treatments are applied to the same experimental units. These designs have certain properties of balance which facilitate the estimation of the effects of treatments and the errors. Designs with 3, 4, 5 and 6 periods and requiring 60 or fewer units for the case in which the effects considered are the direct and residual effects receive most consideration. A number of designs outside this range are included. There are still a number of designs which have not been found or proved non-existent, particularly for non-prime numbers of treatments.

I am indebted to Dr E. M. Patterson for useful suggestions on the mathematics of the section concerned with the designs for prime v of the form $4n+3$, and a minimum value of b/v equal to 2.

REFERENCES

Bose, R. C. (1939). On the construction of balanced incomplete block designs. *Ann. Eugen., Lond.*, **9**, 353.

Bose, R. C. & Nair, K. R. (1941). On complete sets of Latin squares. *Sankhyā*, **5**, 361.

Cochran, W. G., Autrey, K. M. & Cannon, C. Y. (1941). A double change-over design for dairy cattle feeding experiments. *J. Dairy Sci.* **24**, 937.

Fisher, R. A. & Yates, F. (1948). *Statistical Tables for Biological, Agricultural and Medical Research.* Edinburgh: Oliver and Boyd.

Patterson, H. D. (1950). The analysis of change-over trials. *J. Agric. Sci.* **40**, 375.

Patterson, H. D. (1951). Change-over trials. *J.R. Statist. Soc.* (in the Press).

Stevens, W. L. (1939). The completely orthogonalized Latin square. *Ann. Eugen., Lond.*, **9**, 82.

Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Aust. J. Sci. Res.* A, **2**, 149.

Williams, E. J. (1950). Experimental designs balanced for pairs of residual effects. *Aust. J. Sci. Res.* A, **3**, 351.

Yates, F. (1936). Incomplete randomized blocks. *Ann. Eugen., Lond.*, **7**, 121.

MULTI-FACTOR DESIGNS OF FIRST ORDER

G. E. P. BOX

Imperial Chemical Industries, Dyestuffs Division Headquarters, Blackley, Manchester

The problem discussed arises when it is possible to choose in advance the N combinations of levels at which a set of quantitative factors are to be held in a set of N experiments to determine the slopes of a regression surface (assumed planar). It is shown that the minimum variance property of an 'optimum' design arises from the shape of the design pattern and is independent of its orientation. This fact may be utilized as follows:

(1) When prior knowledge of the response surface exists the design may be rotated to reduce possible bias.

(2) The design may be rotated so that systematic effects, such as polynomial time trends and block effects, are eliminated without loss of efficiency.

(3) Subject to the conditions imposed by (2), the orientation of the design may be chosen at random. This has the effect of making the usual normal theory tests exact and completely independent of the distribution of the observations.

1. Suppose that the effect of k quantitative variables or factors $X_1, \ldots, X_i, \ldots, X_k$ (such as time, temperature, concentration) on some measurable response (such as yield of product) is being studied in a region of the response surface that can be represented to a sufficient degree of accuracy by a polynomial equation of degree d. We define a design of order d as an arrangement of experiments which will allow all the coefficients in this polynomial to be separately determined. In this paper it is assumed that d is equal to 1, i.e. that a planar approximation is adequate. It is also assumed that the variables can be controlled exactly at levels decided in advance and that the observed response y differs from η due to experimental error having variance σ^2:

$$E(y) = \eta, \quad E(y-\eta)^2 = \sigma^2. \tag{1}$$

We can perform $N > k$ trials; the problem is to decide which N combinations of levels to use so that the constants defining the plane are estimated with maximum accuracy. We shall not apply the usual limitation that the design is to consist of combinations of a few fixed levels of the factors, but as a means of specifying the extent of variation for a given factor X_i we define the *unit* S_i for this variable as $S_i = \left\{\sum_{u=1}^{N} \frac{(X_{iu}-\bar{X}_i)^2}{N}\right\}^{\frac{1}{2}}$, and write the design in terms of the standardized variables $x_{iu} = (X_{iu}-\bar{X}_i)/S_i$. It will be noted, therefore, that for the standardized variables

$$\sum_{u}^{N} x_{iu} = 0, \tag{2}$$

$$\sum_{u}^{N} x_{iu}^2 = N. \tag{3}$$

A discussion of the problem of scaling and comparing experimental designs will be found in a recent paper (Box & Wilson, 1951), where there is an account of the planning of experiments to attain maxima in connexion with which this investigation was undertaken. The design matrix $\mathbf{D}$ is an $N \times k$ matrix providing a programme of experiments to be performed. The k elements $x_{1u}, \ldots, x_{iu}, \ldots, x_{ku}$ of the uth row are the levels of the standardized variables to be used in the uth trial. They can also be regarded as defining the k co-ordinates of the uth *experimental point* in the k-dimensional factor space. To use the design the

experimenter must decide on suitable average levels $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$ and units $S_1, S_2, \ldots, S_k$ for the variables. The level to be used for the ith variable in the uth trial will then be $X_{iu} = \bar{X}_i + S_i x_{iu}$.

2. Suppose the true regression plane in the region considered is

$$\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k. \tag{4}$$

In an obvious matrix notation the N equations (4) at the N experimental points may be written

$$\boldsymbol{\eta} = \mathbf{X}_1 \boldsymbol{\beta}_1, \tag{5}$$

where $\mathbf{X}_1 = [\mathbf{U} \vdots \mathbf{D}]$ and each element of the column vector $\mathbf{U}$ is unity. If $\mathbf{Y}$ is a column vector of observations $y_1, \ldots, y_u, \ldots, y_N$ made at these experimental points, then providing $\mathbf{X}_1$ is of rank $k+1$ (which implies that $\mathbf{D}$ is of rank k), separate linear estimates $b_0, b_1, \ldots, b_k$ of each of the β's may be calculated. For a particular design $\mathbf{D}$, linear estimates having smallest variances are provided by the method of least squares and are given by $\mathbf{B}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y} = \mathbf{T}_1\mathbf{Y}$, and it is well known that the matrix of variances and covariances for the estimates is $(\mathbf{X}_1'\mathbf{X}_1)^{-1}\sigma^2$. We have then to choose $\mathbf{D}$ so that the diagonal elements of $(\mathbf{X}_1'\mathbf{X}_1)^{-1}$ are minimized.

Consider the symmetrical determinant of sums of squares and products $\Delta = |c_{ij}| = |\mathbf{X}_1'\mathbf{X}_1|$, and denote by C_{hh} the cofactor of c_{hh} in Δ and by $C_{ij.hh}$ the cofactor of c_{ij} in C_{hh}. Using Cauchy's expansion we have

$$\Delta = c_{hh}C_{hh} - Q, \tag{6}$$

where Q is a quadratic form in the k variables c_{hi} $(i = 0, 1, \ldots, h-1, h+1, \ldots, k)$ and

$$Q = \sum_{ij \neq h} c_{hi}c_{hj}C_{ij.hh}. \tag{7}$$

Now since $\mathbf{X}_1$ is of rank $k+1$, Q is necessarily positive definite and Δ is positive. Also $c_{hh} = N$ (from (3)) and $V(b_h) = \sigma^2 C_{hh}/\Delta$, where $V(b_h)$ is the variance of b_h. Consequently, rearranging (6),

$$V(b_h) = N^{-1}\sigma^2\{1 + \Delta^{-1} \text{ (positive definite quadratic form in the } c_{hi})\}.$$

Thus $V(b_h)$ is a minimum only when each of the c_{hi} (the k sums of products of the hth variable with each of the remaining variables) is zero. When this is so all the C_{hi} must be zero also, and consequently b_h is uncorrelated with each of the other estimates and has variance σ^2/N. For maximum efficiency for all the coefficients then $\{c_{ij}\} = N\mathbf{I}$ and a suitable design $\mathbf{D}$ is supplied by any k columns after the first, of a matrix $N^{\frac{1}{2}}\mathbf{O}$, where $\mathbf{O}$ is orthogonal with the elements of its first column all equal. This result was arrived at by Plackett & Burman (1946). They postulated, however, that the x's in their first-order optimum designs should take only the values $+1$ and -1. With this limitation designs existed only when N was a multiple of 4 and they obtained arrangements for $k = 3, 7, 11, \ldots, 99$ factors using $N = 4, 8, 12, \ldots, 100$ trials. Such designs may be used with qualitative or quantitative variables; in our case where the variables are essentially quantitative this restriction is not introduced and N can have any value.

3. For our problem, therefore, designs of optimum precision for up to $k = N-1$ factors in N experiments may be obtained from *any* orthogonal matrix $\mathbf{O}$ with elements in the first column all equal and $\mathbf{X}_1 = N^{\frac{1}{2}}\mathbf{O}$. We now assume $k = N-1$ and consider the geometrical implications of the above result. Since $\mathbf{D}$ is of rank $k = N-1$, the N experimental points are the vertices of an $N-1$ dimensional simplex. Write the uth row of $\mathbf{X}_1$ as $\mathbf{x}_u'$ and denote

the angle which the uth and sth experimental points make with the origin by θ_{us}, then since $\mathbf{X}_1 = [\mathbf{U} \vdots \mathbf{D}]$ the distance of each experimental point from the origin is $(N-1)^{\frac{1}{2}}$ and

$$\mathbf{x}_u' \mathbf{x}_s = 0 = 1 + (N-1)\cos\theta_{us}, \tag{8}$$

i.e.
$$\cos\theta_{us} = -(N-1)^{-1} \quad (\text{all } u \text{ and } s,\ u \neq s). \tag{9}$$

Consequently this design is formed by the vertices of the *regular* $N-1$-dimensional simplex. If two factors are tested in three trials the experimental points should be at the vertices of an equilateral triangle; for three factors tested in four trials the experimental points should form a regular tetrahedron and so on. It should be noted that no restriction is necessary on the orientation of the design. We can turn the regular figure in any direction; this will correspond simply to a different choice of the orthogonal matrix $\mathbf{O}$, and the variance covariance matrix for the b's will remain unchanged.

As an example, Fig. 1 shows two particular orientations of the optimum design for $N = 4, k = 3$.

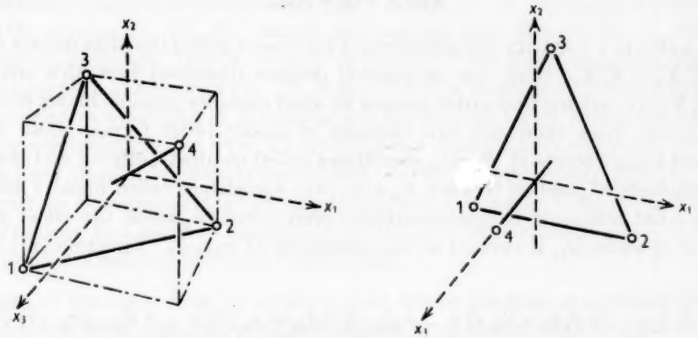

Fig. 1. Orientations of the optimum design for $N=4$, $k=3$.

The design matrices are

$$\mathbf{D}_a = \begin{array}{c} \\ (1) \\ (2) \\ (3) \\ (4) \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \end{array}, \qquad \mathbf{D}_b = \begin{array}{c} \\ (1) \\ (2) \\ (3) \\ (4) \\ \\ (\sqrt{2}) \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ . & 2 & -1 \\ . & . & 3 \end{bmatrix} \\ (\sqrt{2}/\sqrt{3})\ (1/\sqrt{3}) \end{array}. \tag{10}$$

$\mathbf{D}_a$ is the familiar half-replicate of the 2^3 factorial; the other half replicate is obtained by rotation of the first and completes the cube. $\mathbf{D}_b$ is also obtained by orthogonal rotation of $\mathbf{D}_a$, so that the line joining the points (1) and (2) is parallel to the axis of x_1 and (1), (2) and (3) fall on a plane parallel to the plane of x_1 and x_2. $\mathbf{D}_b$ is seen to have elements proportional to Helmert's orthogonal matrix (for clarity the elements are given as whole numbers with the necessary multiplier shown below). This latter design has rather a curious property, for it is a 'one factor at a time design' although not of the orthodox pattern. To use it the experimenter would first perform a 'blank' experiment with all factors at the lower levels; in the second experiment the level of the first factor only would be changed; in all subsequent

experiments this would then be held at the average of these two levels. In the third experiment the level of a second factor would be raised, and in all subsequent experiments this factor would be held at the average level of the three experiments. This procedure could be continued for any number of experiments and factors. The estimates would be uncorrelated, and on the convention we have adopted concerning the units for the factors, the variance of the estimates would be the same with both designs.

4. Although the variances and covariances under orthogonal rotation of the design remain constant, the magnitude and arrangement of the possible biases which might occur if the planar approximation was inadequate do not. Suppose that, contrary to assumption, to obtain a perfect fit it was necessary to include S extra terms $\mathbf{X}_2\boldsymbol{\beta}_2$, so that instead of (5) we had

$$\boldsymbol{\eta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \tag{11}$$

then (Box & Wilson, 1951) $\mathbf{B}_1$ would no longer supply unbiased estimates of $\boldsymbol{\beta}_1$ but instead

$$E(\mathbf{B}_1) = \boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2, \tag{12}$$

where $\mathbf{A}$ is a $(k+1) \times S$ matrix of coefficients of the biases called the alias matrix and given by $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$. With the orthogonal designs discussed here this simplifies to $\mathbf{A} = N^{-1}\mathbf{X}_1'\mathbf{X}_2$. In judging first-order designs we shall consider possible biases due to terms of second order. Now there are two varieties of second-order terms: those which are coefficients of square terms x_1^2, x_2^2, etc., sometimes called quadratic effects, and those which are the coefficients of product terms x_1x_2, x_1x_3, etc., sometimes called linear × linear interactions. In what follows it is mathematically convenient to define the effect β_{11} as the coefficient of x_1^2 whilst β_{12} is defined as the coefficient of $x_1x_2\sqrt{2}$. Equation (11) may then be written

$$\eta = \beta_0 + \beta_1x_1 + \ldots + \beta_kx_k + \beta_{11}x_1^2 + \ldots + \beta_{kk}x_k^2 + \beta_{12}(x_1x_2\sqrt{2}) + \ldots + \beta_{k-1k}(x_{k-1}x_k\sqrt{2}), \tag{13}$$

and the matrices of bias coefficients corresponding to $\mathbf{D}_a$ and $\mathbf{D}_b$ are found to be

$$\mathbf{A}_a = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} 11 & 22 & 33 & 12 & 13 & 23 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & . & . & . \\ . & . & . & . & . & \sqrt{2} \\ . & . & . & . & \sqrt{2} & . \\ . & . & . & \sqrt{2} & . & . \end{bmatrix} \end{array}, \qquad \mathbf{A}_b = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} 11 & 22 & 33 & 12 & 13 & 23 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & . & . & . \\ . & . & . & -2 & -\sqrt{2} & . \\ -\sqrt{2} & \sqrt{2} & . & . & . & -\sqrt{2} \\ -1 & -1 & 2 & . & . & . \end{bmatrix} \end{array} \begin{array}{l} \\ \\ (1/\sqrt{3}) \\ (1/\sqrt{3}) \\ (1/\sqrt{3}) \end{array}. \tag{14}$$

The figures in brackets are multipliers of the rows of $\mathbf{A}_b$. Using $\mathbf{D}_b$, for example, the expected value of b_3 when second-order terms were not all zero would be

$$E(b_3) = \beta_3 - (\beta_{11} + \beta_{22} - 2\beta_{33})/\sqrt{3}.$$

5. If we could, we would choose to orient the optimum design so that the bias coefficients were as small as possible. In this way both random and systematic errors might be simultaneously minimized. Consider the $(k+1) \times (k+1)$ matrix $\mathbf{AA}'$. The sums of squares of bias coefficients for the $k+1$ estimates $b_0, b_1, \ldots, b_k$ are given by the diagonal elements, and the magnitude of these would provide one indication of the efficacy of any particular orientation. We find somewhat unexpectedly, however, that for these optimum first-order designs $\mathbf{AA}'$ is invariant for any orthogonal rotation of the design. This is proved as follows.

Denote by $\mathbf{x}_1', \ldots, \mathbf{x}_u', \ldots, \mathbf{x}_N'$ the N rows of the matrix $\mathbf{X}_1$ and by $\mathbf{X}_3$ a matrix whose N rows are $(\mathbf{x}_1')^{[2]}, \ldots, (\mathbf{x}_u')^{[2]}, \ldots, (\mathbf{x}_N')^{[2]}$ the derived power vectors of degree 2 (Aitken, 1948). Then $\mathbf{X}_3 = [\mathbf{U} \vdots \mathbf{D}\sqrt{2} \vdots \mathbf{X}_2]$ and $N^{-2}\mathbf{X}_1'\mathbf{X}_3\mathbf{X}_3'\mathbf{X}_1 = \mathbf{A}\mathbf{A}' + \mathbf{J}$, where $\mathbf{J}$ is a diagonal matrix in which the first diagonal element is 1 and each of the remaining k diagonal elements is 2. Now suppose the design is submitted to orthogonal rotation and denote the new matrices by $\dot{\mathbf{D}}, \dot{\mathbf{X}}_1, \dot{\mathbf{X}}_3$ and $\dot{\mathbf{A}}$. Then $\dot{\mathbf{D}} = \mathbf{D}\mathbf{G}$, where $\mathbf{G}$ is some $k \times k$ orthogonal matrix, and $\dot{\mathbf{X}}_1 = \mathbf{X}_1\mathbf{H}$, where $\mathbf{H}$ is a $(k+1) \times (k+1)$ orthogonal matrix consisting of $\mathbf{G}$ bordered by a first row $\mathbf{r}' = (1\,0\,0 \ldots 0)$ and a first column $\mathbf{r}$. Now $\mathbf{H}$ transforms the vector $\mathbf{x}_u'$; denote by $\mathbf{H}^{[2]}$ the matrix which correspondingly transforms the vector $(\mathbf{x}_u')^{[2]}$.

Then
$$\dot{\mathbf{A}}\dot{\mathbf{A}}' + \mathbf{J} = N^{-2}\mathbf{H}'\mathbf{X}_1'\mathbf{X}_3\mathbf{H}^{[2]}\mathbf{H}'^{[2]}\mathbf{X}_3'\mathbf{X}_1\mathbf{H}, \tag{15}$$

and since $\mathbf{H}$ is orthogonal so is $\mathbf{H}^{[2]}$. Now the jth diagonal element of $\mathbf{X}_3\mathbf{X}_3'$ is

$$(\mathbf{x}_j')^{[2]}(\mathbf{x}_j)^{[2]} = (\mathbf{x}_j'\mathbf{x}_j)^2 = N^2$$

and the ijth non-diagonal element is $(\mathbf{x}_i')^{[2]}(\mathbf{x}_j)^{[2]} = (\mathbf{x}_i'\mathbf{x}_j)^2 = 0$. Consequently the right-hand side of (15) reduces to $N\mathbf{I}$. We find in consequence that $\mathbf{A}\mathbf{A}' = N\mathbf{I} - \mathbf{J}$, whatever the orientation of the design. That is, the sum of squares of the coefficient of the biases for b_0 is $N-1$, and for each of the effects $b_1 \ldots b_k$ it is $N-2$. The result is of course only true for the particular relative weighting of quadratic and interaction terms which has been adopted. This relative weighting is, however, a reasonable one, and the important conclusion emerges that if we have no prior knowledge concerning the relative importance of particular second-order terms no arrangements which are dramatically worse or better than others can be expected to arise as a result of rotation of the designs. In particular, if $k = N-1$, it is not possible to keep a selected estimate clear of bias.

6. When, on the other hand, something is known of the type of approximating second-degree equation to be expected it might be possible to reduce bias by suitable rotation of the design. Consider a particular class of designs which are such that *only* b_0 is biased by quadratic terms. For this to happen each of the $N-1$ column vectors in $\mathbf{X}_1$ corresponding to first-order effects must have zero inner product with the $N-1$ column vectors in $\mathbf{X}_2$ corresponding to quadratic effects. This can only happen if the latter have all elements equal to $+1$ which in turn implies that elements in $\mathbf{D}$ consist entirely of $+1$'s and -1's. These designs are those obtained by Plackett & Burman. Now the response contours generated by the second degree approximating equation are a set of conics. Suppose the direction of the principal axes of the system were known. Then because of the property mentioned above, if any of the designs of Plackett & Burman were rotated so that their axes were parallel to these principal axes the effects $b_1, \ldots, b_k$ would be unbiased, since in the new variables the second-degree equation contains no product terms. This may have practical application in the exploration of 'ridge' systems. Such systems occur, for example, when a line or plane of near maxima rather than a single point maximum is found. The probable existence and direction of such systems can sometimes be deduced from theoretical considerations, in which case it might be an advantage to rotate axes of the design so as to be parallel to these suspected 'ridges'.

It is worth noting (Box & Wilson, 1951) that by replication of any design with change of signs a design of 'Type B' is obtained. That is to say, one in which the first-order estimates are unbiased by terms of second order. This of course applies to all the designs discussed here.

Elimination of systematic variation

7. If the planar approximation is adequate we have seen that first-order designs of maximum efficiency are obtained by writing down any k mutually orthogonal column vectors each containing N elements and each orthogonal to U, a column vector of unit elements. The latter requirement allows for the elimination of the mean. Now it is easy to show that if any p column vectors are arbitrarily taken we can always write down further $N-p$ column vectors which are mutually orthogonal and each of which is orthogonal to each of the original p vectors. If therefore a first-order design is required for k variables which is such that, not only the mean but also a number of other systematic effects, such as time trends and block effects, are to be eliminated, this may be done by choosing the k orthogonal vectors of the design such that they are orthogonal to the mean and also to a set of $p-1 \leqslant N-k-1$ vectors representing these additional systematic effects.

A field in which there is a need for designs of this sort, and in which the extra trouble in the design and execution of the experiment is amply justified, is that of large-scale plant experiments. Here to avoid the possibility of serious loss of purity or yield, only small changes in the factor levels could usually be tolerated. This would tend to ensure the relative predominance of first-order effects, but a sensitive experiment would be required to detect such effects which would be necessarily small in magnitude. Furthermore, daily alterations to plant conditions would usually not be practicable. On some types of process if a series of trials were to be made it would be necessary to run each for a week at least. This replication would reduce random errors, but long-term trends which are of common occurrence would not be reduced. The usual way of overcoming this difficulty, of course, would be to make comparisons within blocks of weeks, using partial confounding if necessary to reduce block size. An interesting alternative procedure employs the principle discussed above.

Fig. 2 shows a set of fourteen consecutive weekly averages for a certain quality characteristic; it is obvious that a systematic trend with time typical of this particular process is occurring in the individual results. The following analysis is obtained by fitting orthogonal polynomials up to thirteenth order:

Order of term	Sum of squares	Order of term	Sum of squares
1	0·08	11	1·77
2	45·00	12	0·68
3	8·56	13	0·58
4	2·11		
5	0·80	Total	65·25
6	0·87		
7	0·13		
8	4·06	Terms up to fifth order	56·55
9	0·00		
10	0·61	Remainder	8·70

(A useful table giving the orthogonal polynomials for equally spaced ordinates and $N \leqslant 26$ has recently been published by De Lury (1950).)

Analysis of other sets of data from this process covering periods of the same length showed that terms up to fifth order usually accounted for a large proportion of the variation. Suppose now that it was desired to plan a set of fourteen trials on this process each lasting a week, by means of which the first-order effects of three quantitative factors were to be determined with maximum precision and the trend effect eliminated. On the assumption that the 'trend' could be represented by a polynomial of fifth degree plus *independently distributed error terms*, the levels in the 14×3 design matrix could be taken with column elements proportional to any three of the orthogonal polynomials higher than the fifth order or to any three orthogonal linear combinations of these; the remaining five degrees of freedom corresponding to high-order polynomial effects not used up would supply an internal estimate of experimental error appropriate for calculating confidence limits for the effects.

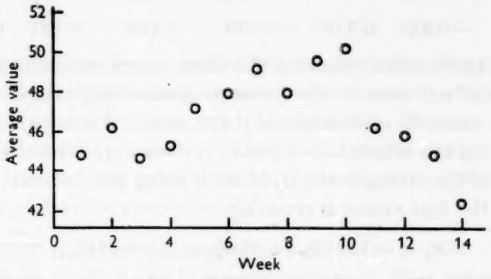

Fig. 2. Fourteen consecutive weekly averages for a quality characteristic.

The assumption that the trend would be represented by a polynomial on time plus independently distributed error terms and consequently that the residual variation would be distributed evenly through the higher order terms would often not be justifiable however. In particular, negative correlation between successive errors which might occur due to incomplete segregation of individual batches could lead to inflated higher order terms. This difficulty may be overcome by the device of randomization. In fact, we choose the directions of the three design vectors in the eight-dimensional residual subspace entirely at random subject only to the restriction that they should be mutually orthogonal. 'Angular' randomization of this sort is achieved conveniently by using a table of random normal deviates such as that of Wold (1948). Any n values of such a table provides a vector $\mathbf{z}_1$ having equal probability of lying in any direction in the n-space. A second vector $\mathbf{z}_{2.1}$ whose direction is random, subject only to the restriction that it is orthogonal to the first, may now be constructed by choosing a second set of n values $\mathbf{z}_2$ from the table and calculating

$$\mathbf{z}_{2.1} = \mathbf{z}_2 - b_{2.1}\mathbf{z}_1, \tag{15}$$

where $b_{2.1} = \mathbf{z}_1'\mathbf{z}_2/\mathbf{z}_1'\mathbf{z}_1$.

In a similar way a third vector $\mathbf{z}_{3.21}$ orthogonal to the first two may be constructed from

$$\mathbf{z}_{3.21} = \mathbf{z}_3 - b_{3.21}\mathbf{z}_{2.1} - b_{3.1}\mathbf{z}_1, \tag{16}$$

where $b_{3.21} = \mathbf{z}_3'\mathbf{z}_{2.1}/\mathbf{z}_{2.1}'\mathbf{z}_{2.1}$ and $b_{3.1} = \mathbf{z}_3'\mathbf{z}_1/\mathbf{z}_1'\mathbf{z}_1$,

and so on. Having obtained $\mathbf{z}_1, \mathbf{z}_{2.1}, \mathbf{z}_{3.21}$, etc., it is convenient to standardize them by dividing the elements of each vector by the square root of the sum of squares of its elements.

These standardized vectors may be denoted by $\mathbf{Z}_1, \mathbf{Z}_{2.1}, \mathbf{Z}_{3.21}$, etc.; their elements are the direction cosines of lines which are mutually perpendicular but whose orientation in the n-space is otherwise random.

In the example quoted the three sets of eight random normal deviates taken from the table were

$$\mathbf{z}_1' = [-0{\cdot}69 \quad 1{\cdot}40 \quad 1{\cdot}79 \quad -0{\cdot}83 \quad 0{\cdot}34 \quad 0{\cdot}19 \quad 1{\cdot}46 \quad 0{\cdot}21],$$
$$\mathbf{z}_2' = [\ 0{\cdot}73 \quad -1{\cdot}00 \quad 0{\cdot}81 \quad 0{\cdot}01 \quad -1{\cdot}13 \quad -1{\cdot}53 \quad -0{\cdot}07 \quad -0{\cdot}22],$$
$$\mathbf{z}_3' = [-0{\cdot}50 \quad -1{\cdot}59 \quad 0{\cdot}82 \quad 0{\cdot}71 \quad 1{\cdot}37 \quad 0{\cdot}90 \quad 0{\cdot}74 \quad -0{\cdot}83].$$

From which were obtained

$$\mathbf{Z}_1' = [-0{\cdot}235 \quad 0{\cdot}476 \quad 0{\cdot}608 \quad -0{\cdot}282 \quad 0{\cdot}116 \quad 0{\cdot}065 \quad 0{\cdot}496 \quad 0{\cdot}071],$$
$$\mathbf{Z}_{2.1}' = [\ 0{\cdot}265 \quad -0{\cdot}332 \quad 0{\cdot}453 \quad -0{\cdot}046 \quad -0{\cdot}453 \quad -0{\cdot}630 \quad 0{\cdot}063 \quad -0{\cdot}080],$$
$$\mathbf{Z}_{3.21}' = [-0{\cdot}133 \quad -0{\cdot}645 \quad 0{\cdot}310 \quad 0{\cdot}270 \quad 0{\cdot}429 \quad 0{\cdot}241 \quad 0{\cdot}241 \quad -0{\cdot}313].$$

These were taken as the direction cosines of the three design vectors in an eight-dimensional subspace whose axes of reference in the fourteen-dimensional sample space were the eight vectors given by the elements of orthogonal polynomials of order 6, 7, ..., 13.

The design vectors in the original co-ordinate system were then obtained by calculating linear combinations of the orthogonal polynomials using the elements of $\mathbf{Z}_1, \mathbf{Z}_{2.1}$ and $\mathbf{Z}_{3.21}$ as coefficients. Thus the first vector is given by

$$\mathbf{x}_1 = -0{\cdot}235\boldsymbol{\xi}_6 + 0{\cdot}476\boldsymbol{\xi}_7 + \ldots + 0{\cdot}071\boldsymbol{\xi}_{13}, \tag{17}$$

where $\boldsymbol{\xi}_s$ is the vector of elements of the sth orthogonal polynomial standardized so that $\boldsymbol{\xi}_s'\boldsymbol{\xi}_s$ 14. The design* thus obtained is given below.

Observation	Factor levels		
	x_1	x_2	x_3
1	−0·2	0·5	0·3
2	0·1	−1·5	−1·1
3	1·2	1·0	1·1
4	−1·6	0·5	0·2
5	−0·4	0·5	0·0
6	0·0	−1·6	−1·5
7	1·6	−0·1	0·5
8	0·6	1·4	0·0
9	−2·0	−0·2	1·4
10	0·1	−1·4	0·2
11	−0·1	1·7	−2·3
12	1·3	−0·7	1·1
13	−0·9	−0·1	0·4
14	0·2	0·1	−0·3

To use the design the three factor levels are varied in proportion to these design elements in suitable units S_i in the manner described in §1. Any trend occurring during the course of the experiment which can be represented by a polynomial of up to fifth degree may be

* Only a single place of decimals is given. More accurate values could of course be used if it were possible to control the levels of the factors more precisely.

eliminated without loss of efficiency. The significance of the effects may be conveniently assessed by means of the analysis of variance:

Source	Degrees of freedom
Due to fifth-order polynomial	5
x_1	1
x_2	1
x_3	1
Residual (error)	5
Total	13

Since the normal theory significance tests depend only on the angles between the vector of observations and the design vectors, angular randomization ensures that these tests are *exact* whatever the nature of the residual variation. The use of these designs is therefore justified whenever the procedure is likely to reduce the residual variation. Because of randomization no assumption need now be made that the trend can be *accurately* represented by the formal mathematical model.

8. In general, when quantitative factors are not restricted to a few fixed levels the scope and flexibility of experimental design is greatly increased. It is hoped to show in a later paper how these ideas may be applied with designs of higher order, and to discuss in more detail the interesting implications of 'angular' randomization.

The drawings were executed by Mr P. S. Ward, to whom I wish to express my best thanks.

REFERENCES

Aitken, A. C. (1948). *Determinants and Matrices*, 5th ed. Edinburgh and London: Oliver and Boyd.
Box, G. E. P. & Wilson, K. B. (1951). *J.R. Statist. Soc.*, Series B, **13**, 1.
De Lury, D. B. (1950). *Values of the Integrals of the Orthogonal Polynomials up to* $n = 26$. University of Toronto Press.
Plackett, R. L. & Burman, J. P. (1946). *Biometrika*, **33**, 305.
Wold, H. (1948). *Tracts for Computors*, no. 25. Cambridge University Press.

TESTS OF SIGNIFICANCE IN CANONICAL ANALYSIS

By F. H. C. MARRIOTT
University of Aberdeen

1. Introduction

Since the earliest papers on multivariate analysis, some twenty years ago, great advances have been made in the techniques used in dealing with multiple measurements. However, there are still considerable difficulties in applying the theory in practical cases. In the simplest case, discrimination between two groups, the problem reduces to one of multiple regression, but in the more general cases of canonical correlation between two sets of variables and discrimination between more than two groups further problems arise which have not yet been completely investigated.

Suppose it is desired to investigate the relationship between two sets of variables $x_1 \dots x_p$ and $y_1 \dots y_q$ measured on $n+1$ individuals. The correlations between linear functions of the x's and y's will be considered; there is consequently a complete duality between the two, and there is no loss of generality in taking $p \leqslant q$. In special cases, either the x's or the y's may be dummy variables corresponding to a division into $p+1$ or $q+1$ groups.

Now the dispersion matrix (the matrix of corrected sums of squares and products) of the y's, say $\mathbf{T}$, with elements T_{ij}, may be broken up into two parts: $\mathbf{W}$ is the dispersion matrix of the y's with the x's eliminated, i.e. the matrix of residual sums of squares and products of the y's after a regression on the x's has been taken out, and $\mathbf{Q} = \mathbf{T} - \mathbf{W}$ the dispersion matrix due to regression. If any linear function of the y's, say $b_i y_i$ (using the summation convention), is considered, an analysis of variance may be carried out on it thus:

	Degrees of freedom	Sum of squares
Regression on the x's	p	$b_i b_j Q_{ij}$
Residual	$n-p$	$b_i b_j W_{ij}$
Total	n	$b_i b_j T_{ij}$

Of the possible vectors $\mathbf{b}$, one gives a maximum value to the variance ratio $\dfrac{b_i b_j Q_{ij}}{b_i b_j W_{ij}} \times \dfrac{n-p}{p}$, and this vector satisfies the q differential equations $\dfrac{\partial}{\partial b_h}\left[\dfrac{b_i b_j Q_{ij}}{b_i b_j T_{ij}}\right] = 0 \,(h = 1, \dots, q)$,

i.e. $$b_k Q_{hk}(b_i b_j T_{ij}) - b_k T_{hk}(b_i b_j Q_{ij}) = 0,$$

or, writing $$l^2 = \frac{b_i b_j Q_{ij}}{b_i b_j T_{ij}},$$

$$\mathbf{b}(\mathbf{Q} - l^2\mathbf{T}) = 0, \quad |\mathbf{Q} - l^2\mathbf{T}| = 0.$$

Now, except in trivial cases, $\mathbf{T}$ is non-singular and $\mathbf{Q}$ is of rank p. Consequently, there are just p distinct non-zero values of l^2, the latent roots of $\mathbf{T}^{-1}\mathbf{Q}$, and p corresponding latent vectors $\mathbf{b}$. The p positive values of l, arranged in descending order of magnitude $l_1 \dots l_p$, are known as the first, second, ..., *canonical correlations* between $\mathbf{x}$ and $\mathbf{y}$, and the corresponding

functions $b_i y_i$ as the first, second, ..., *discriminant functions*. The l's are estimates of the parent canonical correlations $\lambda_1 \dots \lambda_p$. The quantity

$$L = \prod_{i=1}^{p} (1-l_i^2) = \frac{|\mathbf{W}|}{|\mathbf{T}|}$$

has been proposed by Wilks (1932) as a measurement of the degree of association between the x's and the y's. L is small when one or more of the l's is close to 1, and is nearly 1 when there is little association between the x's and y's.

2. TESTS OF SIGNIFICANCE

When $p = 1$, the significance of L may obviously be tested by a variance ratio test. $(1-L)(n-q)/(Lq)$ is distributed as a variance ratio with q and $n-q$ degrees of freedom. The exact test when $p = 2$ was given by Pearson & Wilks (1933). $(1-\sqrt{L})(n-q-1)/(q\sqrt{L})$ is distributed as a variance ratio with $2q$ and $2(n-q-1)$ degrees of freedom.

When $p > 2$, no exact test is available. The approximate test $\chi^2_{[pq]} = -n \log_e L$ is derivable from the theory of likelihood ratio tests, and Bartlett (1938) has obtained from the moments of $\log_e L$ the more accurate test

$$\chi^2_{[pq]} = -\{n - \tfrac{1}{2}(p+q+1)\} \log_e L.$$

This test is sufficiently accurate for most purposes, but Rao (1948) has established a refinement of the test whereby the probability of an observed L is obtained as the sum of an asymptotic series. For all practical purposes, Rao's test may be regarded as an exact test of significance for L.

Further, if $\lambda_1 \neq 0$ and l_1 is an estimate of it, the association between the x's and the y's, apart from the first canonical correlation λ_1, may be measured by $L' = \prod_{i=2}^{p} (1-l_i^2)$. Bartlett (1938) has shown that

$$\chi^2_{[(p-1)(q-1)]} = -\{n - \tfrac{1}{2}(p+q+1)\} \log_e L'$$

provides an approximate test of significance for L'. Similarly, the association can be tested when more than one non-zero canonical correlation has been eliminated.

If the alternative hypothesis to $\lambda_1 \dots \lambda_p = 0$ is $\lambda_1 \neq 0$, $\lambda_2 \dots \lambda_p = 0$, L does not provide an efficient test of independence. A test of significance for l_1, the largest sample canonical correlation, is required. Similarly, if l_1 is obviously significant, a test of significance for l_2 may be required. These tests are not at present available, and it is the purpose of this paper to provide them in some cases

The problem of finding most efficient tests in this instance is complicated by the fact that l_1 does not necessarily correspond to λ_1. The estimated discriminant function associated with l_1 may not be an estimate of the best discriminant function in the population.

3. THE DISTRIBUTION OF THE GREATEST CANONICAL CORRELATION COEFFICIENT

The joint distribution of the canonical correlations in the null case has been given by Fisher (1939) and others. Putting $u_i = l_i^2$, the distribution may be written

$$dF = C \prod_{i=1}^{p} \left\{ u_i^{a-1} (1-u_i)^{b-1} \prod_{j=i+1}^{p} (u_i - u_j)\, du_i \right\},$$

where $a = \tfrac{1}{2}(q-p+1)$, $b = \tfrac{1}{2}(n-q-p+1)$,

$$C = \pi^{\frac{1}{2}p} \prod_{i=1}^{p} \left\{ \frac{\Gamma[\tfrac{1}{2}(n-i+1)]}{\Gamma[\tfrac{1}{2}(q-i+1)]\, \Gamma[\tfrac{1}{2}(n-q-i+1)]\, \Gamma[\tfrac{1}{2}(p-i+1)]} \right\}.$$

Three possible ways of finding a test of significance for l_1 suggest themselves:

(1) By integrating out $l_2 \dots l_p$ the distribution function of l_1 can be obtained. The probability that $u_1 \geqslant U_1$ is given by

$$P = \int_{U_1}^{1} \int_{0}^{u_1} \int_{0}^{u_2} \dots \int_{0}^{u_{p-1}} f(u)\, du_p \dots du_1.$$

In the general case this integral is not easy to evaluate. However, as will be shown later, it is possible to obtain the integral in certain special cases, and in other cases it is possible to derive a significance test, even though the complete integral is not found.

(2) The χ^2 tests for L suggest that a similar approximate test might be derived for l_1. $\chi^2_{(pq)} = -\{n - \frac{1}{2}(p+q+1)\} \log_e (1-l_1^2)$ obviously provides an upper limit for the significance point. Further, consider the approximate break-up of the χ^2 for testing L.

χ^2	Degrees of freedom
$-\{n-\frac{1}{2}(p+q+1)\} \log_e (1-l_1^2)$	$p+q-1$
$-\{n-\frac{1}{2}(p+q+1)\} \log_e \prod_{i=2}^{p} (1-l_i^2)$	$(p-1)(q-1)$
$-\{n-\frac{1}{2}(p+q+1)\} \log_e L$	pq

The second component of χ^2 tests L' and the total χ^2 tests L. The first component does not provide a test for l_1, since the break-up is only valid if $\lambda_1 \neq 0$. If $\lambda_1 = 0$, the selection of the largest l makes the first component a biased test, giving a lower limit for the significance point.

These two tests suggest a possible test

$$\chi^2_{(D)} = -\{n - \tfrac{1}{2}(p+q+1)\} \log_e (1-l_1^2),$$

where D is a symmetrical function of p and q, lying between $p+q-1$ and pq, and therefore equal to q when $p = 1$. Such a test will be discussed in § 6 of this paper.

(3) In view of the difficulty of obtaining the exact distribution, Bartlett (1941) has proposed a rather different test based on the distribution of l_1 given $l_2 \dots l_p$:

$$P = \frac{\int_{U_1}^{1} u_1^{a-1}(1-u_1)^{b-1} \prod_{i=2}^{p} (u_1 - U_i)\, du_1}{\int_{U_2}^{1} u_1^{a-1}(1-u_1)^{b-1} \prod_{i=2}^{p} (u_1 - U_i)\, du_1}.$$

This test may be written in the form

$$P = \frac{\theta(U_1)}{\theta(U_2)},$$

where

$$\theta(U) = B(a+p, b) - B_U(a+p, b) - \sum_{i=2}^{p} U_i \{B(a+p-1, b) - B_U(a+p-1, b)\} + \dots$$

and

$$B_U(a, b) = \int_0^U x^{a-1}(1-x)^{b-1}\, dx.$$

The use of this test is open to various objections. First, the test is almost certainly less powerful, although the fact that l_1 is not necessarily an estimate of λ_1 means that neither is uniformly most powerful. Secondly, this test is tedious and difficult to apply, and cannot be used to tabulate significance levels. Thirdly, if l_2 is close to l_1, the test may not give

significance even to a high value of l_1. This is not such a serious fault as might at first appear; the chief use of the test is when L is not significant, and if two of the l's are large, this will usually show up in the test of L.

4. TESTS OF SIGNIFICANCE FOR l_1*

In certain cases it is possible to obtain the exact integral to test the significance of l_1. For example, if $p = 2$,

$$P = 1 - I_{U_1}(2a, 2b) + \frac{B_{U_1}(a, b)\, U_1^a (1-U_1)^b}{2B(2a, 2b)},$$

where a and b are defined as above, i.e. $a = \frac{1}{2}(q-p+1)$, $b = \frac{1}{2}(n-q-p+1)$, and

$$I_{U_1}(a, b) = \frac{B_{U_1}(a, b)}{B(a, b)}.$$

In particular, when $p = 2$, $q = 3$,

$$P = (2b+1)\, U_1 (1-U_1)^b + (1-U_1)^{2b+1}$$

and when $p = 2$, $q = 5$,

$$P = \tfrac{1}{3}(2b+1)(2b+3)\, U_1^2 (1-U_1)^b + \tfrac{1}{3}(1-U_1)^{2b+1} \{b(2b+1)\, U_1^2 + 3(2b+1)\, U_1 + 3\}.$$

When P is small, i.e. near the significance levels, terms of high order in $1-U_1$ may be neglected and reduced approximate forms are found. These approximate tests should be used only when P is small, but the error in using them near the 5% point is quite negligible (generally of the order of 0·00002 in 0·05, or 0·04%). The approximate tests are

$$p = 2: \quad P \doteqdot \frac{B(a, b)\, U_1^a (1-U_1)^b}{2B(2a, 2b)},$$

$$p = 2, q = 3: \quad P \doteqdot (2b+1)\, U_1 (1-U_1)^b,$$

$$p = 2, q = 5: \quad P \doteqdot \tfrac{1}{3}(2b+1)(2b+3)\, U_1^2 (1-U_1)^b.$$

When $p = 3, q = 4$, the exact test is

$$\begin{aligned} P = {} & (1-U_1)^b \{U_1^2 (b+1)(2b+3) - U_1 (2b+3) + 1\} \\ & + (1-U_1)^{2b+1} \{U_1^2 (b+1)(2b+1) + U_1 (2b+1) + 1\} \\ & - (1-U_1)^{3b+3}. \end{aligned}$$

When P is small, the first term only need be used.

In the general case $p = 3$ the exact test is difficult to obtain, but a similar approximation gives

$$P \doteqdot 1 - I_{U_1}(a, b) + \frac{(2a+2b+1)(a+b)}{ab\, B(a, b)} \left(U_1 - \frac{a}{a+b}\right) U_1^a (1-U_1)^b,$$

which reduces to the first term of the test for $p = 3$, $q = 4$, when $a = 1$.

Finally, for $p = 4$, $q = 5$, the approximate test is

$$P \doteqdot \tfrac{1}{3}(2b+3)(1-U_1)^b \{(b+2)(2b+5)\, U_1^3 - 3(2b+5)\, U_1^2 + 6U_1\}.$$

These tests may be used to test the significance of an observed value of l_1 or to tabulate significance levels of l_1. Further, in the rather common situation when λ_1 is obviously not zero (l_1 being highly significant), the tests may be used to test l_2, reducing p and q by 1. If tables of the incomplete B-function are available the tests are not difficult to apply.

Roy (1945) has recently found the distribution of the largest canonical correlation, and other individual canonical correlations, on the null hypothesis. Roy's distribution is actually

* The derivation of the tests in this section is given in the author's unpublished thesis 'The analysis and interpretation of multiple measurements', University of Aberdeen, 1951.

that of the 'k-statistics', which are related to the canonical correlations by equations of the form

$$t_i^2 = \frac{n-p}{p} \times \frac{l_i^2}{1-l_i^2}.$$

The problem of the distribution of the largest canonical correlation has thus been solved theoretically, but the forms obtained by Roy are not immediately suitable for numerical computation. Tests of significance can more easily be carried out (for $p = 2, p = 3$, and $p = 4, q = 5$) using the formulae of this section.

5. Numerical results

The tabulation of the distribution of l_1 would involve tables of quadruple entry (for p, q, n, and significance level). The significance levels can only be found by a tedious process of trial and error, and it would not be worth while to construct such tables. However, experiments involving multiple measurements are usually only worth doing when a fairly large number of observations can be taken, and the following table of significance levels for large n should prove useful. The table gives the 5% and 1% levels of $\frac{1}{2}nl_1^2$ for n large and $p = 2$, $q = 2(1)12, 21$; $p = 3, q = 3(1)12, 21$; $p = 4, q = 5$.

Significance levels for $\frac{1}{2}nl_1^2$ (n large)

q	$p=2$		q	$p=3$	
	5%	1%		5%	1%
2	4·30	6·08	3	6·56	8·59
3	5·37	7·28	4	7·62	9·76
4	6·34	8·36	5	8·59	10·82
5	7·24	9·37	6	9·54	11·84
6	8·11	10·32	7	10·44	12·82
7	8·94	11·23	8	11·31	13·76
8	9·75	12·12	9	12·16	14·67
9	10·53	12·98	10	12·98	15·56
10	11·30	13·82	11	13·79	16·43
11	12·06	14·64	12	14·59	17·28
12	12·81	15·44	21	21·30	24·45
21	19·15	22·27			

q	$p=4$	
	5%	1%
5	9·81	12·12

Significance levels for intermediate values of q may be found by interpolation. The levels given are reasonably accurate for $n > 100$; for lower values of n, rather lower values of $\frac{1}{2}nl_1^2$ are required for significance.

6. AN APPROXIMATE TEST

Although the tests proposed in § 4 are not difficult to apply, an alternative approximate test appears desirable for two reasons. First, tables of the incomplete B-function are not always available, and secondly, an approximate test which gave good results over the whole range of values at present known might reasonably be used outside this range, in default of any better test.

The possible test $\chi^2_{[D]} = -\{n-\frac{1}{2}(p+q+1)\}\log_e(1-l_1^2)$ has already been mentioned. For $p = 1$ or $q = 1$, $D = p+q-1$, and for higher values of p and q, $p+q-1 < D < pq$. It seems reasonable to take $D = p+q-1+\alpha\{(p-1)(q-1)\}^\beta$. The values of α and β which best fit the percentage points given by the tests in the previous sections are found to be $\alpha = \frac{1}{2}, \beta = \frac{2}{3}$,* so that $D = p+q-1+\frac{1}{2}\{(p-1)(q-1)\}^{\frac{2}{3}}$. D, of course, is not in general integral, and percentage points of χ^2 for fractional degrees of freedom are to be obtained by interpolation in the χ^2 table.

A few percentage points obtained by this test have been calculated for comparison with those in the previous section.

Significance levels for $\frac{1}{2}nl_1^2$ (n large) from the approximate test

q	$p = 2$		q	$p = 3$	
	5%	1%		5%	1%
2	4·33	6·15	3	6·49	8·62
6	8·08	10·41	6	9·37	11·85
12	12·80	15·62	12	14·39	17·33
21	19·24	22·61	21	21·19	24·70

q	$p = 4$	
	5%	1%
5	9·58	12·07

The test provides a good approximation over the range considered. There is some indication that the test tends to overestimate the significance for higher values of p, but with this reservation it seems reasonable to use it outside the range of the tables.

* An additional term in D of the form $\alpha\{(p-1)(q-1)\}^\beta$ was assumed (after various other possibilities had been considered) and the best values of α and β found by a graphical method. The degrees of freedom corresponding to the exact value, d say, were found. $\text{Log}\{d-(p+q-1)\}$ was then plotted against $\log\{(p-1)(q-1)\}$, and α and β estimated. The values, by a fortunate coincidence, were 0·50 and 0·67 respectively. A slight improvement could be obtained by taking the additional term to be of the form $\alpha(p-1)^\beta(q-1)^\gamma$, but the small gain in accuracy seems to be more than offset by the loss of symmetry and simplicity.

As an example of the way in which the test works for small values of n, consider the 5% point for $p = 2, q = 5$. Here $l_1^2 = 1 - \exp\{-14{\cdot}44/(n-4)\}$:

5% points of l_1^2 for $p = 2, q = 5$

n	Exact test	Approximate test
10	0·930	0·910
20	0·600	0·594
50	0·270	0·269
100	0·140	0·140
∞ (nl_1^2)	14·49	14·44

In this case the approximate test gives good results down to $n = 20$. The value $n = 10$ is included only for comparative purposes; it would, of course, be absurd to do a multivariate analysis with such low numbers.

7. Conclusion

In the present paper the exact distribution of the greatest canonical correlation is given for $p = 2$ and $p = 3, q = 4$. Further, a significance test, which may be regarded as exact for all practical purposes, is given for $p = 3$ and $p = 4, q = 5$. 5 and 1% significance levels are given for $p = 2$ and $p = 3$ for selected values of q, and for $p = 4, q = 5$, when n is large.

An approximate test, related to the χ^2 test for Wilks's criterion, is proposed and is shown to be satisfactory for values of p and q for which an exact test is available. The approximate test is $\chi^2_{[D]} = -\{n - \frac{1}{2}(p+q+1)\}\log_e(1-l_1^2)$, where $D = p+q-1+\frac{1}{2}\{(p-1)(q-1)\}^{\frac{1}{2}}$. Outside this range of values there is some reason to suppose the approximate test will give a fairly good result.

When one or more canonical correlations are significant, these tests may be used to test higher canonical correlations. For example, if $\lambda_1 \ldots \lambda_{r-1}$ are not zero, l_r may be tested as if it were the greatest canonical correlation with $p' = p-r+1, q' = q-r+1$.

REFERENCES

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.* **34**, 33.

Bartlett, M. S. (1941). The statistical significance of canonical correlations. *Biometrika*, **32**, 29.

Fisher, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugen., Lond.*, **9**, 238.

Pearson, E. S. & Wilks, S. S. (1933). Methods of statistical analysis appropriate for k samples of two variables. *Biometrika*, **25**, 353.

Rao, C. R. (1948). Tests of significance in multivariate analysis. *Biometrika*, **35**, 58.

Roy, S. N. (1945). The individual sampling distribution of the maximum, the minimum, and any intermediate of the p-statistics on the null hypothesis. *Sankhyā*, **7**, 133.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471.

THE INTERPRETATION OF INTERACTIONS IN FACTORIAL EXPERIMENTS

BY E. J. WILLIAMS

Commonwealth Scientific and Industrial Research Organization, Melbourne

Where the joint effects of two or more factors are not additive, a simple model is proposed for representing the effects. The effects of one factor are assumed to be proportional, rather than equal, at different levels of the other factors.

The main effects of the first factor are given as weighted averages of the simple effects at the different levels of the other factors, the weights being the estimated factors of proportionality. The weights are given as the latent vector of a matrix of sums of squares and products corresponding to the largest latent root of the matrix; the sum of squares for the weighted main effect is a multiple of this latent root, and the other latent roots correspond to a partition of the interaction sum of squares.

The analysis is closely related to the canonical analysis of a set of variates.

Tests of significance of (*a*) the residual interactions and (*b*) the adequacy of a proposed set of weights are discussed.

For the case where the matrix has only two non-vanishing latent roots, the approach of the joint distribution of the roots to its limiting form is discussed. The joint probability density is expanded as a series of Bessel functions of imaginary argument. Asymptotic formulae for the moments and product-moments of the roots are derived.

Exact tests for the adequacy of a proposed set of weights, when there are only two non-vanishing latent roots, are presented.

The methods of analysis are illustrated with a numerical example.

I. INTRODUCTION

In a factorial experiment the treatments whose effects are compared comprise all possible combinations of two or more factors, each of which is applied at two or more different levels, in contrast to experiments in which only one factor is varied, the others being held constant. The results of a factorial experiment provide information, not only about the average effects of each factor, but also as to the manner in which the effects of one factor vary as other factors are changed. The results of such an experiment are most easily interpreted when the effects of the different factors are additive; that is, when the effects of one factor are independent of variations in other factors. When the effects of different factors are not additive, attention must be paid to the way in which they interact. The purpose of this note is to present a method which has been found useful in interpreting interaction effects. It is based on a simple assumption about the mode of interaction of a pair of factors.

Where the effects of two or more factors are not additive, it seems reasonable to assume that the effects of one factor (as measured, for example, by the set of differences among the results for different levels of the factor), rather than being *constant*, are *proportional* at different levels of the other factors. When there are two factors, A and B, and the A effects are assumed proportional in this way, at different levels of B, the expected effect of the combination of level i of A with level j of B may be expressed in the form

$$a_i c_j + b_j.$$

The 'main effects' for the different levels of A would then be represented by the constants a_i, which are determinate apart from a constant factor; the constants c_j may be regarded as weights to be applied to the results at different levels of B, in determining the A effects. In general, the c_j may be either positive or negative. The effects of B would be given by the constants b_j. It is noted that this representation of the treatment effects is not symmetrical

with respect to the factors A and B. This kind of effect often seems to occur; for instance, when A represents different types of fertilizer, and B represents different levels of application, in an agricultural experiment.

If this assumption is consistent with the data obtained from an experiment, interaction effects found to exist may be satisfactorily accounted for by the methods of analysis set out below. R. A. Fisher in 1935 (see Fisher, 1949, §50) applied such a method in the analysis of an experiment on quantity and type of nitrogenous fertilizer. The effects of fertilizer type are assumed to be proportional to quantity applied, an assumption which is found to fit the facts. It is not necessary to the method, however, even where one factor is quantitative, to assume the effects of the other factor to be proportional to quantity, though such an assumption, if it proves to be realistic, is certainly to be preferred. In many experiments of the type described by Fisher, the effects will not be proportional to quantity of fertilizer, but will increase less rapidly than quantity. In general, the weights for the effects of factor A at different levels of B are to be estimated from the data, being so chosen as to maximize the sum of squares for the 'main effect' of A, and hence to minimize the residual interaction sum of squares. If the interactions thus determined prove to be non-significant, the interpretation of the data may be regarded as satisfactory.

The earliest use of a method of this kind for analysing a factorial experiment appears to be due to Fisher & Mackenzie (1923), in the analysis of a varietal and manurial trial with potatoes. It was assumed that the expected yield of any variety-manure combination was the product of two constants, representing the variety and the manurial level. It was there shown that the sums of squares for main effects and interactions are given by the latent roots of a matrix of sums of squares and products. The present problem is also closely allied to that of comparison of different scales of measurement, as discussed by Cochran (1943). In fact, the same basic methods are used, though there are some additional preliminaries, such as the determination of the effects of factor B, the testing of the significance of the unweighted interactions, and checking the adequacy of weights chosen *a priori*.

Testing the significance of interactions involves the same problems as arise in testing departures from collinearity of several populations in discriminant function analysis. While the distribution theory involved in such tests is far from complete, the available tests appear to be quite adequate for practical purposes. The problem of testing significance will be discussed briefly, particular attention being given to the simpler case in which the matrix has only two latent roots. The method will be illustrated with a practical example.

II. Determination of main effects to minimize interactions

For simplicity we consider an experiment with two factors, A and B, applied at n and p levels respectively, and replicated r times. We write

x_{ij} mean for level i of A in combination with level j of B,
$x_{i.}$ mean for level i of A over all levels of B,
$x_{.j}$ mean for level j of B over all levels of A.

Without loss of generality we may regard the mean of all results as zero.

In accordance with the model set out above, the expected value of x_{ij} may be expressed as

$$a_i c_j \theta + b_j,$$

where

$$\Sigma a_i = \Sigma b_j = 0. \tag{1}$$

The introduction of the constant θ, whose significance will appear later, enables a_i and c_j to be chosen to satisfy the further conditions

$$\Sigma a_i^2 = \Sigma c_j^2 = 1. \tag{2}$$

The method of least squares then gives the following equations for the constants; treating the $a_i\theta$ as one set of constants and minimizing

$$\sum_i \sum_j (x_{ij} - a_i c_j \theta - b_j)^2 + 2\alpha\Sigma a_i\theta + 2\beta\Sigma b_j + \gamma\Sigma c_j^2,$$

we have

$$\sum_j c_j(x_{ij} - a_i c_j \theta - b_j) - \alpha = 0, \tag{3}$$

$$\theta \sum_i a_i(x_{ij} - a_i c_j \theta - b_j) - \gamma c_j = 0, \tag{4}$$

$$\sum_i (x_{ij} - a_i c_j \theta - b_j) - \beta = 0. \tag{5}$$

Use being made of the relations (1) and (2), these equations reduce to

$$a_i\theta + \sum_j b_j c_j = \sum_j c_j x_{ij}, \tag{6}$$

$$c_j\theta = \sum_i a_i x_{ij}, \tag{7}$$

$$b_j = x_{.j}, \tag{8}$$

and from (6) and (8),

$$a_i\theta = \sum_j c_j(x_{ij} - x_{.j}). \tag{9}$$

From (7) and (9) we may eliminate either the a_i or the c_j; it is preferable, as will be seen, to eliminate the set of variates with the greater number of values. We shall assume that n exceeds p, and eliminate the a_i:

$$c_j\theta^2 = \sum_i \sum_k c_k(x_{ik} - x_{.k})x_{ij}. \tag{10}$$

If

$$t_{jk} = \sum_i (x_{ij} - x_{.j})(x_{ik} - x_{.k})$$

$$= \sum_i x_{ij}x_{ik} - nx_{.j}x_{.k}, \tag{11}$$

θ^2 is seen to be a latent root of the matrix T whose typical element is t_{jk}. This matrix is of order $p \times p$, and of rank p; in general, its rank is the lesser of $n-1$, p.

Had the c been eliminated, θ^2 would have been found as a latent root of the matrix U, where

$$u_{hi} = \sum_j (x_{hj} - x_{.j})(x_{ij} - x_{.j}). \tag{12}$$

This matrix, while having the same non-zero latent roots as T, is of order $n \times n$ and consequently has $n-p$ irrelevant zero roots if n exceeds p, and one zero root otherwise.

The sum of the latent roots of either matrix, multiplied by r, is

$$r\sum_i \sum_j (x_{ij} - x_{.j})^2 = r(\sum_i \sum_j x_{ij}^2 - n\sum_j x_{.j}^2),$$

that is, the sum of squares for the main effect of A and for the interactions of A and B. If

$$r\theta^2 = s,$$

then the largest value s_1 of s is the maximum value which the weighted sum of squares between treatments can take. Accordingly, the corresponding weights are those which minimize the residual sum of squares $(s_2 + s_3 + \ldots)$. This weighting may thus be regarded as

giving the 'main effect' of factor A, and s_1 as the sum of squares for this weighted main effect. The other values of s, namely, $s_2, s_3, \ldots$, are seen to be orthogonal components of the sum of squares for residual interaction effects.

If only the value s_1 is significant, and the values $s_2, s_3, \ldots$ are non-significant, then the interactions have been successfully accounted for. However, even where s_1 alone does not account for all the significant effects, the fact that s_2 is the maximum component of the residual interactions, s_3 the maximum of the remainder after eliminating the effect corresponding to s_2, and so on, enables the effects corresponding to the largest components of interaction to be identified, and separately tested for significance.

Corresponding to any latent root, the values of a_i and c_j may be determined. Since, in practice, only the values corresponding to the largest root are required, the simplest method of calculation is the iterative method. This may be based on the use of equation (10) or a similar equation for the a_i. Alternatively, equations (7) and (9) may be used alternately to derive successive approximations for the a_i and c_j. Equation (7) may be written as

$$c_j\theta = \Sigma a_i(x_{ij} - x_{.j}),$$

so that the calculations can be performed directly on a table of the values $x_{ij} - x_{.j}$ (or the corresponding totals). This method obviates the calculation of the matrix T, but the convergence of the iteration is only about half as rapid as that using equation (10). Either of these methods simultaneously yields successive approximations to the largest latent root, as well as the constants.

III. Determination of interaction components when unweighted main effects are required, and in other cases

Sometimes the unweighted main effects for each factor are relevant, though such a situation does not seem often to arise. In such a case the interactions, defined in the ordinary way, may be separated into components representing effects of decreasing magnitude. This enables an interpretation of any significant interactions to be made. As the actual procedures, both algebraic and computational, do not differ in principle from those used in the case discussed above, only the results need be presented.

The expected value of x_{ij} is

$$c_i d_j \theta + a_i + b_j,$$

where
$$\Sigma a_i = \Sigma b_j = \Sigma c_i = \Sigma d_j = 0, \tag{1'}$$

and
$$\Sigma c_i^2 = \Sigma d_j^2 = 1. \tag{2'}$$

The equations for the c_i and d_j are

$$c_i\theta = \sum_j d_j(x_{ij} - x_{.j}), \tag{13}$$

$$d_j\theta = \sum_i c_i(x_{ij} - x_{i.}). \tag{14}$$

Values of θ^2 are found as latent roots of the matrix V, where

$$\begin{aligned} v_{jk} &= \sum_i (x_{ij} - x_{i.} - x_{.j})(x_{ik} - x_{i.} - x_{.k}) \\ &= \sum_i (x_{ij} - x_{i.})(x_{ik} - x_{i.}) - n x_{.j} x_{.k}. \end{aligned} \tag{15}$$

The quantities $s_1, s_2, s_3, \ldots$ are the canonical components of the interaction sum of squares, whose significance may be separately tested.

Another alternative, of more practical importance, is to weight both main effects. This method derives from the assumption that the expected effect (either absolute, or as a deviation from the general mean) is the product of constants representing the effects of the two factors. Such an assumption was made by Fisher & Mackenzie (1923).

The sums of squares for main effects and interactions are then given by the latent roots of the matrix W, where

$$w_{jk} = \sum_i x_{ij} x_{ik}.$$

Since these different cases are all treated in the same way, and the first described seems most important in practice, these others will not be considered further.

IV. Extension to data with partially confounded interactions

The method of analysis outlined above applies where the experiment has been so arranged that all treatment comparisons are of equal accuracy (treatment effects being either unconfounded, or all confounded to the same extent). Assuming no interaction effects, we find the matrix of expectations of the elements t_{jk} in this case to be a multiple of I, the unit matrix. The analysis leads to the solutions of the matrix equation

$$|T - \theta^2 I| = 0.$$

In cases where the (unweighted) interactions have been partially confounded, the method of analysis has to be somewhat modified. For simplicity, we consider only the case where all such interactions have been confounded to the same extent. We suppose that the interaction effects are estimated with an efficiency factor E; that is, the variance of interaction comparisons is $1/E$ times that of main effects. For a $n \times n$ factorial experiment, using blocks of n experimental units, where it is possible to confound all the interaction effects equally in a multiple of $n-1$ replications, without confounding any of the main effects, we have

$$E = \frac{n-2}{n-1}.$$

In the analysis, use is made of the estimates of the effects of one factor at each level of the other (which, following Yates, we shall call simple effects). These estimates reduce to the mean x_{ij} only when there is no confounding; their calculation in other cases is standard procedure (see, for example, Yates, 1937), and will not be discussed here. In determining the weighting which maximizes the sum of squares for weighted main effects, we find that quantities t_{jk} are involved, defined as above, except that they are, in the general case, functions of the simple effects and not of the treatment means. The matrix of expectations of the t_{jk} is now proportional to J, where

$$EJ = \begin{bmatrix} 1-e & -e & -e & \dots & -e \\ -e & 1-e & -e & \dots & -e \\ -e & -e & 1-e & \dots & -e \\ \dots & \dots & \dots & \dots & \dots \\ -e & -e & -e & \dots & 1-e \end{bmatrix}$$

and

$$e = \frac{1-E}{n}.$$

From this result it follows that, corresponding to any set of weights w_j, the sum of squares for weighted main effects is

$$r\sum_j\sum_k w_j w_k t_{jk},$$

provided

$$\frac{1}{E}\{\Sigma w_j^2 - e(\Sigma w_j)^2\} = 1.$$

We therefore choose the scale of the weights to satisfy this condition.

Moreover, it can be shown that two sets of weights, w_j and w'_j, correspond to independent comparisons if and only if

$$\Sigma w_j w'_j - e(\Sigma w_j)(\Sigma w'_j) = 0,$$

and that consequently the correlation between the comparisons corresponding to the two sets of weights will be

$$\frac{1}{E}\{\Sigma w_j w'_j - e(\Sigma w_j)(\Sigma w'_j)\}.$$

The solution of the normal equations leads to the matrix equation

$$|T - \theta^2 J| = 0;$$

and the sums of squares for the weighted main effects and for the canonical components of interaction are given by the values of $r\theta^2$.

A typical example is the 3×3 factorial, laid out in blocks of three units. With two replications, it is possible to confound all the interaction comparisons by half. Then $E = \frac{1}{2}$, and we have

$$J = \tfrac{1}{3}\begin{bmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{bmatrix}.$$

Such a case provides the basis for the numerical example.

V. One factor at two levels only

When factor A is at two levels only, the matrix T has only one latent root. The weights for any level of factor B are proportional to the observed differences between two levels of A. The weighted main effect accounts for all the effects of factor A, and interactions are annihilated. The simplicity of this case arises from the fact that the equations of estimation of all the constants are linear; consequently, the distribution theory and significance tests required are those occurring in the analysis of variance.

In particular, it is a simple matter to test whether the weights observed differ significantly from any set of weights chosen *a priori*, and thus, provided the weights are sufficiently accurately determined, to derive fiducial limits at any given level of probability for the true relative magnitudes of the A effects at different levels of B.

Suppose that the assigned weights are w_j, scaled so that

$$\Sigma w_j^2 = 1.$$

Then if

$$y_j = x_{1j} - x_{2j},$$

the sum of squares for the weighted A effect is

$$\tfrac{1}{2}r(\Sigma w_j y_j)^2.$$

The remainder, or interaction, is $\tfrac{1}{2}r\{\Sigma y_j^2 - (\Sigma w_j y_j)^2\}$,

a sum of squares with $p-1$ degrees of freedom, which may be tested for significance against the error mean square. Sets of values of w_j which render the remainder significant may then be rejected as discordant with the data. No such rejection is possible unless $\frac{1}{2}r\Sigma y_j^2$, treated as a sum of squares with $p-1$ degrees of freedom, is significant.

VI. TESTS OF SIGNIFICANCE

The tests of significance relevant to the results of the analysis of interaction effects will consist of tests of the adequacy of the assumptions made in carrying out the analysis. The basic assumption is that, with main effects suitably defined, interactions do not exist. This is equivalent to assuming that, if

$$\phi = \{\phi_{ij}\}$$

is the matrix of expected effects of factor A, then ϕ is of rank 1; that is, the matrix

$$\Phi = r\phi'\phi$$

has only one non-zero latent root. We shall denote this latent root by λ.

In this case, the weights c_j would be proportional to the set of elements of any row of the effect matrix ϕ. If these weights were known, the significance of the interactions (or, what is equivalent, the existence of other non-zero roots of Φ) could be tested by means of an analysis of variance. The sum of squares for the weighted main effect would then be proportional to a non-central χ^2, with parameter λ, and the residual interactions would be distributed independently of λ.

When the weights are not known but estimated from the data, the result of sampling errors in the weights is that all the sample latent roots are correlated, and have distributions depending on the population latent root. Some allowance must then be made for the effect of λ on the sample roots representing interactions. It will be shown that, provided the population latent root is sufficiently large, a fairly accurate test for other roots can be made, based on the observed interactions. Such a test can be seen to be equivalent to a 'test for collinearity', as described by Fisher (1938).

Another aspect of the data which should be tested is the adequacy of any system of weights chosen *a priori*, or, in general, fiducial limits for the true weights. Here again, a fairly accurate test is possible, provided the population root is large. Fisher (1940) has discussed such tests as applied to discriminant functions.

Since the method of analysis described is of no interest unless effects of treatment factor A are known to exist, a test of the main effect is clearly not relevant. However, the fact that such effects exist is likely to ensure that the population latent root is large, so that significance tests are more satisfactory than they might otherwise be.

We shall assume for convenience that the general variance in the population is unity. If the population root λ is large, the distribution of s_1 approximates a non-central χ^2 distribution, with $n+p-2$ degrees of freedom and parameter λ. Any proposed set of weights will lead to a sum of squares, s_w for the weighted main effect of factor A which, by hypothesis, will have the same parameter λ, and $n-1$ degrees of freedom. Hence, in this case, the adequacy of any proposed set of weights can be tested by the difference $s_1 - s_w$, a central χ^2 with $p-1$ degrees of freedom.

In the same way, since the sum of squares for A effects and for interactions is a non-central χ^2 with $(n-1)p$ degrees of freedom, the sum of squares for interactions may be regarded as having a χ^2 distribution with $(n-2)(p-1)$ degrees of freedom, and may be tested for significance accordingly.

VII. Theoretical basis of the tests applied when λ is large, and discussion of the approach of the distribution to its limiting form

The case of λ large can be seen to give the distributions stated above if we consider the population 'main effect'. The correlation between the observed main effect and the population main effect will be near to unity, when λ is large, and since the interaction components are orthogonal to the observed main effect, their correlation with the population effect will be nearly zero. In deriving the observed main effect, $p-1$ independent constants are fitted, so that the effect has in all $n+p-2$ degrees of freedom. The interaction sum of squares will then have $(n-2)(p-1)$ degrees of freedom, and be virtually independent of λ. The tests given then follow. (This argument can of course be extended to the case where two or more population roots are large, in deriving the limiting distribution of the sum of the remaining sample roots.)

It is instructive to consider the approach of the joint distribution to the limiting form, and this will now be done for the case where there are only two sample roots which are not identically zero. It is hoped to generalize the results given here in later communications. For this case, let $p = 2$. The joint probability density of s_1 and s_2, when $\lambda = 0$, is (see, for example, Hsu, 1939)

$$\frac{1}{4(n-3)!}\, e^{-\frac{1}{2}(s_1+s_2)}(s_1 s_2)^{\frac{1}{2}(n-4)}(s_1-s_2)\, ds_1 ds_2. \tag{16}$$

The population root introduces a factor (see Bartlett, 1947)

$$e^{-\frac{1}{2}\lambda}\sum_{i=0}^{\infty}\frac{(\frac{1}{2}\lambda)^i}{i!}\frac{m_i}{\mu_i}, \tag{17}$$

where

$$m_i = \frac{1}{\pi}\sum_{k=0}^{i}\frac{\Gamma(i-k+\frac{1}{2})\,\Gamma(k+\frac{1}{2})}{\Gamma(i-k+1)\,\Gamma(k+1)}\, s_1^{i-k}s_2^k,$$

$$\mu_i = \frac{2^i\Gamma[\frac{1}{2}(n-1)+i]}{\Gamma[\frac{1}{2}(n-1)]}.$$

Clearly, when $\lambda = 0$, $\qquad E(m_i) = \mu_i$.

The limiting distribution is (Hsu, 1940)

$$\frac{1}{2^{n-1}\,\Gamma[\frac{1}{2}(n-2)]}\, e^{-\frac{1}{2}(\lambda+s_1+s_2)}s_1^{\frac{1}{2}(n-2)}s_2^{\frac{1}{2}(n-4)}\sum_{i=0}^{\infty}\frac{(\frac{1}{4}\lambda s_1)^i}{i!\,\Gamma(\frac{1}{2}n+i)}\, ds_1 ds_2$$

$$= \frac{1}{\Gamma[\frac{1}{2}(n-2)]}\, e^{-\frac{1}{2}(\lambda+s_1+s_2)}\left(\frac{s_1}{\lambda}\right)^{\frac{1}{2}(n-2)}\left(\frac{s_2}{2}\right)^{\frac{1}{2}(n-4)} I_{\frac{1}{2}(n-2)}(\sqrt{(\lambda s_1)})\frac{ds_1}{2}\frac{ds_2}{2}, \tag{18}$$

s_1 and s_2 now being independent.

The joint distribution function can conveniently be expressed as an asymptotic series, with leading term equal to the limiting distribution function, the series consisting of terms involving Bessel functions of increasing order. In deriving the series, it is to be remembered that, while

$$s_1 = O(\lambda),$$

$$s_2 = O(1).$$

Put $\lambda s_1 = u_1^2, \quad \lambda s_2 = u_2^2;$

then the joint density is

$$\frac{\exp\left[-\frac{1}{2}\left(\lambda + \frac{u_1^2+u_2^2}{\lambda}\right)\right]}{2^{\frac{1}{2}(n-4)}\Gamma\frac{1}{2}(n-2)\,\lambda^{n-1}} u_1^{\frac{1}{2}n} u_2^{n-3}\left(1-\frac{u_2^2}{u_1^2}\right)\sum_{i=0}^{\infty}\frac{(\frac{1}{2}u_1)^{\frac{1}{2}(n-2)+2i}\,\Gamma(\frac{1}{2})}{i!\,\Gamma[\frac{1}{2}(n-1)+i]}\left(\frac{m_i}{s_1^i}\right) \tag{19}$$

and

$$\frac{m_i}{s_1^i} = \frac{1}{\pi}\sum_{k=0}^{\infty}\frac{\Gamma(i-k+\frac{1}{2})\,\Gamma(k+\frac{1}{2})}{\Gamma(i-k+1)\,\Gamma(k+1)}\left(\frac{u_2}{u_1}\right)^{2k}.$$

We now sum the series in (19), with respect to i for given values of k. In illustration of the method, we give the summation for $k = 0$. The sum is

$$\sum_{i=0}^{\infty}\frac{\left(\frac{u_1}{2}\right)^{\frac{1}{2}(n-2)+2i}}{i!\,\Gamma[\frac{1}{2}(n-1)+i]}\frac{\Gamma(i+\frac{1}{2})}{\Gamma(i+1)} = \sum_{i=0}^{\infty}\frac{\left(\frac{u_1}{2}\right)^{\frac{1}{2}(n-2)+2i}}{i!\,\Gamma(\frac{1}{2}n+i)}\frac{\Gamma(i+\frac{1}{2})\,\Gamma(\frac{1}{2}n+i)}{\Gamma(i+1)\,\Gamma[\frac{1}{2}(n-1)+i]}$$

$$= \sum_{i=0}^{\infty}\frac{\left(\frac{u_1}{2}\right)^{\frac{1}{2}(n-2)+2i}}{i!\,\Gamma(\frac{1}{2}n+i)}F(\tfrac{1}{2},\tfrac{1}{2}(n-2);\ \tfrac{1}{2}n+i;\ 1)$$

$$= \sum_{i=0}^{\infty}\frac{\left(\frac{u_1}{2}\right)^{\frac{1}{2}(n-2)+2i}}{i!\,\Gamma(\frac{1}{2}n+i)}\left(1+\frac{\frac{1}{4}(n-2)}{1!\,(\frac{1}{2}n+i)}+\frac{(\frac{1}{4})^2 1.3.(n-2)\,n}{2!\,(\frac{1}{2}n+i)(\frac{1}{2}n+i+1)}+\ldots\right)$$

$$= I_{\frac{1}{2}(n-2)}(u_1)+\frac{1.(n-2)}{1!\,2u_1}I_{\frac{1}{2}n}(u_1)+\frac{1.3.(n-2)\,n}{2!\,(2u_1)^2}I_{\frac{1}{2}(n+2)}(u_1)+\ldots. \tag{20}$$

On summing the series for all values of k, and collecting terms of the same order, we have for the summed terms in (19),

$$I_{\frac{1}{2}(n-2)}+\left|+\frac{n-2}{2u_1}I_{\frac{1}{2}n}+\frac{u_2^2}{2u_1^2}I_{\frac{1}{2}(n-2)}+\right|+\frac{3(n-2)\,n}{8u_1^2}I_{\frac{1}{2}(n+2)}+\frac{nu_2^2}{4u_1^3}I_{\frac{1}{2}n}+\frac{3u_2^4}{8u_1^4}I_{\frac{1}{2}(n-2)}+\left|\right.$$

$$+\frac{5(n-2)\,n(n+2)}{16u_1^3}I_{\frac{1}{2}(n+4)}+\frac{3n(n+2)\,u_2^2}{16u_1^4}I_{\frac{1}{2}(n+2)}+\frac{3(n+2)\,u_2^4}{16u_1^5}I_{\frac{1}{2}n}+\frac{5u_2^6}{16u_1^6}I_{\frac{1}{2}(n-2)}+\left|\right.+\text{etc.}, \tag{21}$$

where the argument of the Bessel functions is omitted for convenience. The numerical coefficients in each set of terms are seen to be the same as those occurring in the m_i, namely,

$$\frac{1}{\Pi}\frac{\Gamma(i-k+\frac{1}{2})\,\Gamma(k+\frac{1}{2})}{\Gamma(i-k+1)\,\Gamma(k+1)}.$$

The series (21), substituted for the sum in (19), gives the required representation of the joint probability density with terms of decreasing order of magnitude in λ. From this representation, the joint moments of s_1 and s_2 may be determined as asymptotic series in λ.

In the calculation of moments, and for other purposes, it can be assumed that the ranges of integration of s_1 and s_2 (and hence also of u_1 and u_2) are independent. The errors thus introduced decrease exponentially with λ, and consequently do not enter into the asymptotic formulae.

The general term in the expansion of the probability density is of the form

$$k\exp\left[-\tfrac{1}{2}\left(\lambda+\frac{u_1^2+u_2^2}{\lambda}\right)\right]u_1^{p-1}u_2^{q-1}I_l(u_1), \tag{22}$$

and similar terms will appear in the integrand in the determination of any product-moment.

The integral of such an expression (see, for example, Watson, 1944, p. 393) is

$$k.2^{\frac{1}{2}(p+q-t)-2}\lambda^{\frac{1}{2}(p+q+t)}\frac{\Gamma[\frac{1}{2}(p+t)]\,\Gamma(\frac{1}{2}q)}{\Gamma(t+1)}\,{}_1F_1(\tfrac{1}{2}(t-p)+1;\ t+1;\ -\tfrac{1}{2}\lambda), \tag{23}$$

where ${}_1F_1$ is a generalized hypergeometric function (Watson, 1944, p. 100).

The function $${}_1F_1(e;\ f;\ -\tfrac{1}{2}\lambda) \quad (e<f)$$

has the asymptotic expansion

$$\frac{\Gamma(f)}{\Gamma(f-e)}\left\{\frac{2^e\Gamma(e)}{\lambda^e\Gamma(f-e)}-\frac{2^{e+1}\Gamma(e+1)}{\lambda^{e+1}\Gamma(f-e-1)}+\frac{2^{e+2}\Gamma(e+2)}{2!\,\lambda^{e+2}\Gamma(f-e-2)}\cdots\right\}, \tag{24}$$

as may be seen by considering the differential equation which it satisfies. Actually it is easier to consider the function

$$e^{\frac{1}{2}\lambda}\,{}_1F_1(e;\ f;\ -\tfrac{1}{2}\lambda) = {}_1F_1(f-e;\ f;\ \tfrac{1}{2}\lambda),$$

which leads to equivalent results.

The product-moments may now be derived.

$$\begin{aligned}E(u_1^{2g}u_2^{2h}) = (2\lambda)^{g+h}\Bigg(&\frac{\Gamma(\frac{1}{2}n+g)\,\Gamma[\frac{1}{2}(n-2)+h]}{\Gamma(\frac{1}{2}n)\,\Gamma[\frac{1}{2}(n-2)]}\,{}_1F_1(-g;\ \tfrac{1}{2}n;\ -\tfrac{1}{2}\lambda)+\\ &+\frac{1}{2}\left(\frac{n-2}{n}\right)\frac{\Gamma(\frac{1}{2}n+g)\,\Gamma[\frac{1}{2}(n-2)+h]}{\Gamma(\frac{1}{2}n)\,\Gamma[\frac{1}{2}(n-2)]}\,{}_1F_1(-g+1;\ \tfrac{1}{2}(n+2);\ -\tfrac{1}{2}\lambda)\\ &-\frac{1}{2}\frac{\Gamma[\frac{1}{2}(n-2)+g]\,\Gamma(\frac{1}{2}n+h)}{\Gamma(\frac{1}{2}n)\,\Gamma[\frac{1}{2}(n-2)]}\,{}_1F_1(-g+1;\ \tfrac{1}{2}n;\ -\tfrac{1}{2}\lambda)+\Bigg),\end{aligned} \tag{25}$$

with similar terms which may be derived in an obvious way. On replacing the ${}_1F_1$ by their asymptotic expansions, we find, after some reduction, and replacing the u by their values in terms of the s,

$$\begin{aligned}E(s_1^g s_2^h) = \lambda^g(n-2)_h\Bigg\{&1+\frac{2g^2+g(n-2)-h}{\lambda}\\ &+\frac{1}{\lambda^2}\left[2g^4+2g^3(n-4)+\frac{g^2}{2}(n^2-10n+20)-\frac{g}{2}(n-2)(n-6)-2g^2h-gh(n-6)-\frac{3h}{2}-\frac{h^2}{2}\right]\\ &+\frac{1}{\lambda^3}[\tfrac{4}{3}g^6+2g^5(n-6)+\tfrac{1}{3}g^4(3n^2-42n+124)+\tfrac{1}{6}g^3(n^3-30n^2+224n-432)\\ &-\tfrac{1}{3}g^2(3n^3-54n^2+288n-416)+\tfrac{1}{3}g(n^3-15n^2+74n-96)-2g^4h-2g^3h(n-8)\\ &-\tfrac{1}{2}g^2h(n^2-18n+82)+\tfrac{1}{2}gh(n^2-15n+70)-h(n+3)-g^2h^2-\tfrac{1}{2}gh^2(n-10)-\tfrac{9}{2}h^2-\tfrac{1}{2}h^3]+\ldots\Bigg\},\end{aligned} \tag{26}$$

where $$(n-2)_h = (n-2)\,n(n+2)\ldots(n+2h-4).$$

The expressions for the moments are somewhat complicated, and it is advisable to make use of the numerous checks which exist. Any symmetric polynomial in s_1 and s_2 has as expectation a polynomial in λ, of degree equal to the highest power (say f) of s_1 occurring. Consequently, in the asymptotic expansion of the expectation, all terms beyond the $(f+1)$th vanish. These results provide the required checks, which may be enumerated as follows:

Terms in expansion	Check polynomials
2	1
3	s_1+s_2, s_1s_2
4	$s_1^2+s_2^2$, $s_1^2s_2+s_1s_2^2$, $s_1^2s_2^2$
⋮	

Explicitly, we have

$$E(s_1+s_2) = \lambda+2(n-1),$$
$$E(s_1 s_2) = (n-2)[\lambda+(n-1)],$$
$$E(s_1^2+s_2^2) = \lambda^2+2\lambda(n+2)+2(n-1)(n+2),$$
$$E(s_1^2 s_2+s_1 s_2^2) = (n-2)[\lambda^2+3\lambda(n+1)+2(n-1)(n+1)],$$
$$E(s_1^2 s_2^2) = (n-2)n[\lambda^2+2\lambda(n+1)+(n-1)(n+1)].$$

(a) Test for significance of interaction

Since s_2 is not distributed independently of λ, it would be desirable to find some function of s_1 and s_2, distributed independently of λ, which would provide a fully satisfactory test for the significance of interactions. Such a function does not appear to exist. From the form of the limiting distribution for large λ it can be seen that the function would be equivalent to s_2 at this limit. Below is given a function whose mean is independent of λ for λ large, and tends to s_2, but its other moments are functions of λ. It therefore appears that only an approximate test is ever possible.

The moments of s_2 are

$$E(s_2^h) = (n-2)_h\left\{1-\frac{h}{\lambda}-\frac{h^2+3h}{2\lambda^2}-\frac{h^3+9h^2+2h(n+3)}{2\lambda^3}\ldots\right\}.$$

By comparing moments, it can be seen that

$$s_2\left(1+\frac{1}{\lambda}+\frac{n+3}{\lambda^2}\right)$$

has approximately a χ^2 distribution with $n-2+\dfrac{n(n-2)}{\lambda^2}$ degrees of freedom.

In practice, λ being unknown, we may use

$$s_2+\frac{s_2^2}{ns_1}+\frac{s_2^3}{(n+2)s_1^2},$$

whose mean is $$n-2+O(\lambda^{-3}),$$

and variance $$(n-2)n-\frac{4(n-2)}{\lambda}+\frac{2(n-2)(n+1)(n+4)}{n\lambda^2}+O(\lambda^{-3}).$$

The means, variances and covariance of s_1 and s_2 are also of interest. We have

$$E(s_1) = \lambda+n+(n-2)\left(\frac{1}{\lambda}+\frac{2}{\lambda^2}+\frac{n+8}{\lambda^3}+\ldots\right),$$
$$E(s_2) = (n-2)\left(1-\frac{1}{\lambda}-\frac{2}{\lambda^2}-\frac{n+8}{\lambda^3}-\ldots\right),$$
$$V(s_1) = 4\lambda+2n-\frac{4(n-2)}{\lambda}-\frac{2(n-2)(n+7)}{\lambda^2}-\frac{4(n-2)(7n+22)}{\lambda^3}-\ldots,$$
$$V(s_2) = 2(n-2)-\frac{4(n-2)}{\lambda}-\frac{2(n-2)(n+3)}{\lambda^2}-\frac{4(n-2)(5n+6)}{\lambda^3}-\ldots,$$
$$\operatorname{cov}(s_1,s_2) = (n-2)\left(\frac{4}{\lambda}+\frac{2(n+5)}{\lambda^2}+\frac{8(3n+7)}{\lambda^3}+\ldots\right).$$

The correlation between s_1 and s_2 is accordingly

$$\sqrt{\left(\frac{2(n-2)}{\lambda^3}\right)}+O(\lambda^{-\frac{1}{2}}),$$

and can clearly be ignored if λ is large, and n small; indeed, since λ must be $O(n^{\frac{1}{2}})$ to be significant, as is shown below, a significant value of λ implies a correlation of at most

$$O(\lambda^{-\frac{1}{2}})$$

for large n.

(*b*) *Large values of n*

Another instructive limiting case is that of large n. While not of practical importance for the type of data discussed here, it does give some indication of the way in which s_1 and s_2 depend on λ, while proving more tractable than the general case.

If we put

$$s_1 = n+y_1\sqrt{(2n)},$$
$$s_2 = n+y_2\sqrt{(2n)},$$
$$\lambda = \quad \eta\sqrt{(2n)},$$

the limiting distribution is found to be

$$\frac{1}{2\sqrt{\pi}}\exp\{-\tfrac{1}{2}[y_1^2+y_2^2-\eta(y_1+y_2)+\eta^2]\}(y_1-y_2)\,I_0[\tfrac{1}{2}\eta(y_1-y_2)]\,dy_1dy_2.$$

The derivation shows, incidentally, that, to be shown as significant, λ must be of order $n^{\frac{1}{2}}$. The interesting feature of this distribution is that, as can readily be seen by a change of variables, y_1+y_2 and y_1-y_2 are distributed independently, though both are dependent on the parameter η. When we put

$$z_1 = y_1+y_2, \quad z_2 = y_1-y_2,$$

the joint distribution is

$$\frac{1}{2\sqrt{\pi}}e^{-\frac{1}{4}(z_1-\eta)^2}dz_1\,\tfrac{1}{2}e^{-\frac{1}{4}(z_2^2+\eta^2)}I_0\left(\frac{z_2\eta}{2}\right)z_2\,dz_2.$$

The sum z_1 has a normal distribution with mean η and variance 2; z_2 has a distribution which, when η is zero, becomes that of $\sqrt{(2\chi^2)}$ for two degrees of freedom, and which, for large η, may be put in the asymptotic form

$$\frac{1}{2\sqrt{\pi}}e^{-\frac{1}{4}(z_2-\eta)^2}\left(\frac{z_2}{\eta}\right)^{\frac{1}{2}}dz_2.$$

The mean value of z_2 can best be found from the series

$$\eta+\frac{1}{\eta}+\frac{1}{2\eta^3}+\frac{3}{2\eta^5}+\ldots$$

and the variance from

$$2-\frac{2}{\eta^2}-\frac{4}{\eta^4}-\frac{22}{\eta^6}-\ldots.$$

It is seen that, when η is large, z_1 and z_2 are distributed with approximately the same mean and variance, so that y_2 becomes irrelevant for estimating η.

VIII. TEST FOR ADEQUACY OF ASSIGNED WEIGHTS

Corresponding to any set of weights w_j for the effects of factor A at the jth level of factor B, there will be a sum of squares s_w for the weighted main effect of A. Clearly, any criterion for testing the concordance of these weights with the experimental results will involve the difference $s_1 - s_w$. A criterion of this kind can, of course, only be applied after it has been shown that interactions have been accounted for by the weighting applied; or, in other words, that there is no evidence for the existence of more than one population latent root, as indicated by the sample roots $s_2, s_3, \ldots$.

When $p = 2$ and there are only two roots s_1 and s_2, any value of s_w must lie between them. The simultaneous distribution of these three quantities may be found from the following considerations. A test of the weights w_j is actually a test of the hypothesis that the non-zero population root λ corresponds to just these weights. Under these conditions, s_w is seen to be a sufficient statistic for λ. Its distribution is, in fact, that of a non-central χ^2 with $n-1$ degrees of freedom and parameter λ. The simultaneous *conditional* distribution of s_1 and s_2, for fixed s_w, is therefore independent of λ, depending only on the given value of s_w.

The simultaneous distribution of s_1, s_2 and s_w is found to be

$$\frac{1}{4(n-3)!} e^{-\frac{1}{2}(s_1+s_2)} (s_1 s_2)^{\frac{1}{2}(n-4)} (s_1 - s_2)\, ds_1 ds_2 \times \frac{e^{-\frac{1}{2}\lambda}}{\pi} \frac{ds_w}{\sqrt{[(s_1 - s_w)(s_w - s_2)]}} \left(1 + \frac{\frac{1}{2}\lambda s_w}{n-1} + \frac{(\frac{1}{2}\lambda s_w)^2}{2!\,(n-1)(n+1)} + \ldots\right),$$

so that the conditional distribution of s_1 and s_2 is

$$\frac{1}{2^{\frac{1}{2}(n-1)} \sqrt{\pi}\, \Gamma[\frac{1}{2}(n-2)]} e^{-\frac{1}{2}(s_1+s_2-s_w)} \frac{(s_1 s_2)^{\frac{1}{2}(n-4)}}{s_w^{\frac{1}{2}(n-3)}} \frac{(s_1 - s_2)}{\sqrt{[(s_1 - s_w)(s_w - s_2)]}} ds_1 ds_2.$$

We now make a change of variables, putting

$$v_1 = \frac{(s_1 - s_w)(s_w - s_2)}{s_w},$$

$$v_2 = \frac{s_1 s_2}{s_w}.$$

The simultaneous probability density of v_1 and v_2 is found to be

$$\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}v_1} v_1^{-\frac{1}{2}} dv_1 \times \frac{1}{2^{\frac{1}{2}(n-2)} \Gamma[\frac{1}{2}(n-2)]} e^{-\frac{1}{2}v_2} v_2^{\frac{1}{2}(n-4)} dv_2,$$

so that v_1 is distributed as χ^2 with 1 degree of freedom, while v_2 is independently distributed as χ^2 with $n-2$ degrees of freedom.

While v_1 and v_2 formally are symmetric functions of s_1 and s_2, it can be seen that, from the circumstance that s_w will, for acceptable hypotheses, lie in the neighbourhood of s_1, v_1 will depend mainly on s_1 and approximate to $s_1 - s_w$, while v_2 will depend mainly on s_2. Thus, v_2 provides a test of the residual interaction effects, while v_1 provides an independent test of the adequacy of the chosen set of weights.

It will be noted that the test provided by v_2, while exact, is somewhat 'fluid', in that it depends on the choice of s_w. In practice this does not cause any difficulty; any set of weights which makes v_1 or v_2 significant may be regarded as unacceptable. (An adjustment can be made to the significance levels used for each test, to ensure that chance rejection due to

either cause is of the selected frequency. If the significance probability level is α, we may take probability levels α_1 and α_2 for the two tests, such that $\alpha_1+\alpha_2-\alpha_1\alpha_2=\alpha$, since the tests are independent.)

The lower bound for v_2 is s_2, corresponding to the choice of $s_w=s_1$. The systematic use of s_2, regarded as having a χ^2 distribution with $n-2$ degrees of freedom, results in underestimation of significance, as has already been indicated above (§VII (*a*)).

The criterion v_1 may be used to determine fiducial limits for s_w, and hence for the corresponding weights; however, it will be noted that v_1 is not a monotonic function of s_w, as it has a maximum value

$$(\sqrt{s_1}-\sqrt{s_2})^2,$$

when

$$s_w=\sqrt{(s_1 s_2)}.$$

Hence, for sufficiently high probability levels no possible value of s_w will be found discordant with the data. This is as would be expected.

If we have an independent estimate of error, a mean square M based on q degrees of freedom, and F is the tabular value of the F-distribution for 1 and q degrees of freedom, and if we consider only departures of main effects and not of interactions then the fiducial limit for s_w is given by the equation

$$v_1=\frac{(s_1-s_w)(s_w-s_2)}{s_w}=FM,$$

whence

$$s_w=\tfrac{1}{2}\{s_1+s_2-FM+\sqrt{[(s_1-s_2)^2-2FM(s_1+s_2)+F^2M^2]}\},$$

the larger root being the relevant one.

When $n=3$, there are likewise two roots s_1 and s_2, but since sets of weights can be chosen giving a sum of squares less than s_2, the foregoing analysis does not apply. There are then, in fact, two canonical comparisons, corresponding to s_1 and s_2, together with a set of $p-2$ comparisons corresponding to zero roots. We can, however, derive a valid test for assigned weights and for interactions by considering the regression of the w_j on the c_j.

Let ${}_1c_j$, ${}_2c_j$ be the sets of weights corresponding to s_1, s_2 respectively, and put

$$r_1=\sum_j {}_1c_j w_j, \quad r_2=\sum_j {}_2c_j w_j.$$

Then

$$s_w=r_1^2 s_1+r_2^2 s_2;$$

and if we put

$$s_r=\frac{s_w}{r_1^2+r_2^2},$$

we have

$$s_1>s_r>s_2.$$

It can be shown that

$$v_1=\frac{(s_1-s_r)(s_r-s_2)}{s_r}$$

is distributed as χ^2 with 1 degree of freedom, while

$$v_2=\frac{s_1 s_2}{s_r}$$

is distributed as χ^2 with $p-1$ degrees of freedom. Also

$$v_3=s_r-s_w$$

is likewise distributed with $p-2$ degrees of freedom. The test of the chosen weights is therefore given by v_1+v_3, with $p-1$ degrees of freedom, while v_2 gives the test for interactions.

These tests will be discussed more fully in later communications, where it is also intended to give generalizations to the case of three or more sample latent roots.

IX. NUMERICAL EXAMPLE

The data for the following example are provided by an experiment on the effects of time of contact of catalyst (A), and concentration of catalyst (B), on the dry tensile strength of a plastic material. Each factor is at three levels, and there are two replications of each treatment combination. The replications were carried out over six days, in such a way that all the interaction effects were one-half confounded with differences between days, while main effects remained unconfounded.

Table 1. *Experimental results (with level of factor A in brackets)*

Level of B ...		1	2	3	Total
Replication	Day				
1	1	11·61 (1)	11·94 (2)	12·21 (3)	35·76
	2	11·78 (2)	11·98 (3)	11·94 (1)	35·70
	3	11·75 (3)	12·72 (1)	10·82 (2)	35·29
2	4	12·89 (1)	12·38 (3)	11·94 (2)	37·21
	5	11·99 (3)	12·17 (2)	11·80 (1)	35·96
	6	11·57 (2)	12·39 (1)	12·11 (3)	36·07
Total		71·59	73·58	70·82	215·99

The experimental results are set out in Table 1, and the analysis of variance, performed without weighting, in Table 2. In the latter, sums of squares and mean squares have been multiplied by 18 to avoid rounding-off errors, and this factor is retained throughout the calculations. The analysis shows that interactions of the two factors are significant. Further interpretation is based on the assumption that the effects of time of contact (A) are proportional at different levels of concentration (B).

The estimates of the simple effects, multiplied by 18, are presented in Table 3. To obtain the value for the combination of the first level of each factor, for instance, the results of Table 1 are multiplied by the following set of factors:

$$\begin{array}{rrr} 8 & -4 & -4 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \\ 8 & -4 & -4 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{array}$$

and similarly for the other simple effects.

From Table 3 the values of $324t_{jk}$ are readily derived as sums of squares and products of the deviations of the values in different rows from their means. The quantities $36t_{jk}$ are set out in Table 4, together with the matrix J.

The sum of squares for each canonical comparison is given by twice the corresponding root of the equation

$$|T - \theta^2 J| = 0.$$

Consequently, each root ϕ of $|36T - \phi J| = 0$ will correspond to a sum of squares multiplied by $\frac{36}{2} = 18$, which is the factor already being

used in our analyses of variance. The equation for ϕ is a cubic, which may be written explicitly as

$$4\phi^3 - 136{\cdot}7512\phi^2 + 228{\cdot}87889056\phi = 0.$$

The roots of this equation, and the corresponding sets of weights, are given in Table 5. From

Table 2. *Analysis of variance* (*a*), *main effects unweighted*

	Degrees of freedom	Sum of squares (× 18)	Mean square (× 18)
Blocks	5	12·7337	2·5467
Main effects: *A*	2	12·1706	6·0853*
B	2	15·5018	7·7509*
Interaction, *AB*	4	22·0172	5·5043*
Error	4	2·0832	0·5208
Total	17	64·5065	

* Significant at the 5 % level.

Table 3. *Estimates of simple effects and main effects* (× 18)

Level of *B*	Level of *A*			Mean
	1	2	3	
1	0·34	15·01	− 3·17	4·06
2	− 3·77	4·24	− 16·46	− 5·33
3	− 0·23	− 5·00	9·04	1·27
Mean	− 1·22	4·75	− 3·53	0·00

Table 4. *Sums of squares and products* (× 36) *of estimates of A effects at different levels of B*

The matrix J

$$\begin{matrix} 20{\cdot}6682 & 19{\cdot}9398 & -13{\cdot}2504 \\ 19{\cdot}9398 & 24{\cdot}2106 & -16{\cdot}5360 \\ -13{\cdot}2504 & -16{\cdot}5360 & 11{\cdot}3262 \end{matrix} \quad \tfrac{1}{3}\begin{bmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{bmatrix}$$

Table 5. *Latent vectors and latent roots*

Level of *B* (*j*)	Latent vectors			Latent roots from Table 4
	1	2	3	
${}_1c_j$	0·55872	0·59747	− 0·11208	32·4230
${}_2c_j$	0·59255	− 0·23243	0·45275	1·7648
${}_3c_j$	− 0·05818	0·50565	0·67016	0·0000

Table 6. *Analysis of variance* (*b*) *weighted main effects of factor A*

	Degrees of freedom	Sum of squares (× 18)	Mean square (× 18)
Main effects: *A*	4	32·4230	8·1058*
Interaction: *AB*	2	1·7648	0·8824 (*n*)
Sum: $A + AB$	6	34·1878	

* Significant at the 5 % level.
(*n*) Not significant.

Table 7. *Analysis of variance* (*c*) *test of assigned weights*

	Degrees of freedom	Sum of squares (× 18)	Mean square (× 18)
Assigned weights	2	31·7844	15·8922
Departures from assigned weights	2	0·6183	0·3092 (*n*)
Interactions	2	1·7851	0·8925 (*n*)
Sum: $A + AB$	6	34·1878	

(*n*) Not significant.

the weighted analysis of variance, Table 6, we see that the weighting satisfactorily accounts for the interaction effects.

In practice it will be convenient to adopt a set of weights which is simpler than that given by the largest value of θ^2, and it will then be desirable to test the concordance of these weights with the data. For the present example, the set

$$w_1 = w_2 = \sqrt{\tfrac{3}{8}}, \quad w_3 = 0$$

suggests itself.

With the notation of § VIII, and using the results of § IV, we have

$$\begin{aligned} 18s_w &= \tfrac{3}{8}(20{\cdot}6682 + 2 \times 19{\cdot}9398 + 24{\cdot}2106) \\ &= 31{\cdot}7844. \end{aligned}$$

$$\begin{aligned} r_1 &= 2\sqrt{\tfrac{3}{8}}[{}_1c_1 + {}_1c_2 - \tfrac{1}{3}({}_1c_1 + {}_1c_2 + {}_1c_3)] \\ &= 0{\cdot}98977; \end{aligned}$$

similarly,

$$r_2 = 0{\cdot}10921,$$

$$r_3 = 0{\cdot}09176.$$

Note, as a check, that

$$r_1^2 + r_2^2 + r_3^2 = 1.$$

It follows that

$$\begin{aligned} 18s_r &= \frac{18s_w}{r_1^2 + r_2^2} \\ &= 32{\cdot}0452. \end{aligned}$$

Then

$$\begin{aligned} 18v_1 &= 18\,\frac{(s_1 - s_r)(s_r - s_2)}{s_r} \\ &= 0{\cdot}3484, \end{aligned}$$

$$\begin{aligned} 18v_2 &= 18\,\frac{s_1 s_2}{s_r} \\ &= 1{\cdot}7851. \end{aligned}$$

The analysis of variance for testing the adequacy of the proposed weights is set out in Table 7. The usual significance tests show that the weights are satisfactory.

REFERENCES

BARTLETT, M. S. (1947). Multivariate analysis. *Suppl. J.R. Statist. Soc.* **9**, 176.

COCHRAN, W. G. (1943). The comparison of different scales of measurement for experimental results. *Ann. Math. Statist.* **14**, 205.

FISHER, R. A. (1938). The statistical utilization of multiple measurements. *Ann. Eugen., Lond.*, **8**, 376.

FISHER, R. A. (1940). The precision of discriminant functions. *Ann. Eugen., Lond.*, **10**, 422.

FISHER, R. A. (1949). *The Design of Experiments.* Edinburgh: Oliver and Boyd.

FISHER, R. A. & MACKENZIE, W. A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. *J. Agric. Sci.* **13**, 311.

HSU, P. L. (1939). On the distribution of roots of certain determinantal equations. *Ann. Eugen., Lond.*, **9**, 250.

HSU, P. L. (1940). On generalized analysis of variance. I. *Biometrika*, **31**, 221.

WATSON, G. N. (1944). *A Treatise on the Theory of Bessel Functions.* Cambridge University Press.

YATES, F. (1937). The design and analysis of factorial experiments. *Tech. Commun. Bur. Soil Sci., Harpenden*, no. 35.

ON SAMPLING FROM A POPULATION OF RANKERS

By A. S. C. EHRENBERG

Statistical Laboratory, University of Cambridge

1. It is clear from the recent symposium on ranking methods (Moran, Whitfield & Daniels, 1950) and Kendall's book (1948) that relatively little is known of the theory of sampling from a population of 'rankers' or finite rankings—the problem of m rankings—particularly under anything but what may be called the 'null-hypothesis' of random ranking. A few remarks, even if rather incomplete, on measures of agreement among the rankings may therefore be of interest.

2. Consider m rankings of a set of n objects. Of a number of rather similar measures of agreement, the popular one nowadays is the coefficient of concordance W (Kendall & Babington Smith, 1939; Kendall, 1948), which is defined as the sum of squares of the n totals of the m rank values for each object about the mean $\frac{1}{2}m(n+1)$, divided by the maximum of this expression, $m^2(n^3-n)/12$, which occurs if all m rankings agree. Here each ranking is denoted by the numbers 1 to n. The null-hypothesis that the ranks are given at random can be tested by using the fact that under it $(m-1)W/(1-W)$ is distributed approximately as the variance-ratio F-distribution with $\nu_1 = n-1-2/m$ and $\nu_2 = (m-1)\nu_1$ degrees of freedom. Exact tables of the distribution when m and n are both small are given by Kendall (1948), who also discusses the effects of 'ties'. A related test for randomness, proposed by Friedman (1937), is that $m(n-1)W$ is distributed as χ^2 with $(n-1)$ degrees of freedom, if $m(n-1) > 7$. Stuart (1951) discusses the distribution of W more generally.

The coefficient of concordance W is a generalization of Spearman's well-known rank correlation coefficient ρ, which is defined in terms of the squared differences between the two rankings (represented again by the numbers 1 to n) of each object, i.e.

$$\rho = 1 - \frac{6}{n^3-n} \sum_{i=1}^{n} (\text{1st ranking} - \text{2nd ranking of } i\text{th object})^2.$$

In fact (Kendall, 1948) if $\rho_{\text{av.}}$ stands for the mean of the coefficients ρ between the ${}_mC_2$ pairs of rankings, then

$$\rho_{\text{av.}} = \frac{mW-1}{m-1}.$$

Now, whatever is felt in general about the relative merits as measures of rank correlation of Spearman's and other coefficients such as Kendall's τ (see below), in this case the Spearman approach does not seem so appropriate. In it, the ranking of any two objects can be too much affected by what has happened to the other objects, which matters particularly if the objects to be ranked are arbitrarily selected. For example, if six out of seven rankers ($m = 7$) rank a certain object, i, highest, and all seven rank another object, j, second, then i is ranked higher than j significantly more often at about the 5 % probability level. Yet if the seventh person gives object i some rank bigger than 8, then j would be ranked higher than i according to the '$\rho - W$' kind of reckoning; what may make this more absurd is that it could happen only if there are actually more than eight objects to be ranked.

3. Some coefficient based on the number r_{ij} of the m rankers who agree on ordering object i higher than j $(i, j = 1, \ldots, n)$ seems more attractive than W. It can easily be seen that about the simplest function of the r_{ij} to handle is $r_{ij}(m - r_{ij})$. An attractive property of this, compared with any possible *linear* function such as max. $(r_{ij}, m - r_{ij})$, is that strong agreement about any one pair of objects is weighted heavily. Constructing a coefficient u which takes the value 1 for complete agreement and whose minimum value is roughly zero, one obtains

$$u = 1 - \frac{2}{MN} \sum_{i<j}^{n} r_{ij}(m - r_{ij}),$$

where $M = {}_mC_2$, $N = {}_nC_2$. This is, in fact, the average of the ${}_mC_2$ 'Kendall' rank correlation coefficients τ between the m rankings taken in pairs, where τ is defined by the same formula, putting $m = 2$. The precise minimum value of u is $-1/(m-1)$ for m even and $-1/m$ for m odd, and its expected value for random rankings is 0. Kendall (1948) claims for W that it ranges from 0 to 1, and a similar coefficient might be preferred here, i.e. one related to u in the same way that W is related to $\rho_{\text{av.}}$. However, W can actually only take the value of 0 if $m(n+1)$ is even, and the coefficients u and $\rho_{\text{av.}}$ seem slightly more attractive measures since for each there are *two* points whose meaning is unaltered by changes in m and n. It is important to note that u is simple to calculate.

4. It is, perhaps, not surprising that the coefficient u turns out to be identical with the coefficient of agreement u for paired comparisons defined by Kendall & Babington Smith (1940; Kendall, 1948), who, however, considered it as 'not of much practical value' for ranked data. In the method of paired comparisons, n objects are ranked in pairs by each of the m observers, there being $N = {}_nC_2$ pairs in all. It is clear that a person may make comparisons inconsistent with a one-dimensional ranking of the n objects, and it is one of the attractions of the method that this can be tested (e.g. Kendall, 1948).

For the null-hypothesis that the comparisons have been made at random, it has been shown (e.g. Kendall, 1948) that the distribution of u can be approximated to by a χ^2 form as follows:

$$\chi^2 = \frac{2N}{(m-2)}(u+1) - (m-3)\nu,$$

where $\nu = 2NM/(m-2)^2$ is the appropriate number of degrees of freedom. As the number of objects n increases, the distribution tends to normality with variance $1/MN$.

For the non-null case, suppose that the probability of ranking the ith object higher than the jth is π_{ij}, estimated by r_{ij}/m. (It is unnecessary to specify any relations between the π's.) The expected value of u is then given by

$$E(u) = 1 - \frac{4}{N} \sum_{i<j}^{n} \pi_{ij}(1 - \pi_{ij}).$$

Further, considering the first few moments,* it is easy to show that the distribution of u tends to normality for either large m or n, with variance

$$\text{var.}(u) = \frac{8}{MN^2} \sum_{i<j}^{n} \{(m-1)[\pi_{ij}(1-\pi_{ij}) - 4\pi_{ij}^2(1-\pi_{ij})^2] + 2\pi_{ij}^2(1-\pi_{ij})^2\},$$

except that for n fixed the distribution is a χ^2 under the null-hypothesis of all π's $= \frac{1}{2}$, as already mentioned. This compares with the distribution of the multiple correlation

* Mr B. Babington Smith has kindly drawn my attention to similar formulae described by him in the discussion of the paper by Professor A. S. C. Ross (1950: *J.R. Statist. Soc.* B, **12**, 54).

coefficient R when the population value is or is not equal to zero (Fisher, 1928). It may be noted that var. (u) in general can be larger than the value $1/MN$ in the null-case.

5. As is to be expected, the distribution of u from ranked data, where in each ranking the comparisons of pairs of objects cannot be independent, is not so simple. But certain results come easily enough. First one should notice that it follows at once from the well-known result for the correlation of ρ and τ under the hypothesis of random ranking (e.g. Kendall, 1948), that the correlation of u and W is

$$\frac{2(n+1)}{\sqrt{(4n^2+10n)}},$$

which tends to 1 very rapidly as n increases. (Clearly all coefficients which are at all reasonable should agree in this case.) But for non-random rankings the values of the two coefficients can differ; for example, for 'circular' rankings ($m = n$), e.g.

123

231

312,

$W = 0$ and $u = 1 - 4/m \to 1$ as m increases (cf. §2).

Kendall & Babington Smith (1940) give, as 'a matter of theoretical interest', a χ^2 approximation for the sampling distribution of u on the hypothesis of random ranking. Since the result there contained an error, no doubt merely a misprint, it seems worth while to derive it anew. The first few moments of u are

$$E(u) = 0,$$

$$E(u^2) = \frac{(2n+5)}{9MN},$$

$$E(u^3) = \frac{2(m-2)(2n^2+6n+7)}{27M^2N^2},$$

$$E(u^4) = \frac{2MN(2n+5)^2}{54} - \frac{(6n^3+21n^2+31n+31)}{225} + \frac{(m-2)(m-3)(2n^3+8n^2+12n+9)}{27M^3N^3},$$

where again $M = {}_mC_2, N = {}_nC_2$. These expressions are obtained by generalizing Daniels's (1944) method for deriving the sampling variance of τ; it should be noted that the account of this in Kendall's book (1948) seems to contain a number of false steps, besides some confusion (cf. Moran, 1951) due to taking over the notation of Daniels's original paper.

The Pearsonian coefficients β_1 and β_2 are of the order $8/n$ and $(3+12/n)$ respectively, indicating a type III distribution, which turns out to be

$$\chi^2 = \frac{6(2n+5)MN}{(m-2)(2n^2+6n+7)}u + \nu,$$

with

$$\nu = \frac{2(2n+5)^3 MN}{(m-2)^2(2n^2+6n+7)^2} \text{ degrees of freedom.}$$

For large n the distribution of u tends to normality with variance $(2n+5)/(9MN)$.

To test the goodness of fit of the χ^2 approximation the exact distribution of u has been calculated for some small values of m and n; the labour of this increases very rapidly with

m and n. For three objects and four or five rankers, and for four objects and three rankers, we have frequencies f as follows:

$n = 3, m = 4$	$18u$	18, 12, 10, 6, 4, 2, 0, −2, −4, −6;
	f	1, 8, 6, 20, 24, 6, 28, 60, 48, 15;
$n = 3, m = 5$	$30u$	30, 22, 18, 14, 10, 6, 2, −2, −6;
	f	1, 10, 30, 30, 100, 95, 240, 420, 370;
$n = 4, m = 3$	$18u$	18, 14, 10, 6, 2, −2, −6;
	f	1, 9, 33, 82, 135, 165, 151.

Applying a continuity correction of taking the mean of the calculated u and the next smallest value (for large m, n, subtract $1/MN$ for m even, $2/MN$ for m odd), one obtains the following probabilities for exceeding certain values in percentages:

	$n = 3, m = 4$	$n = 3, m = 5$	$n = 4, m = 3$
Exact probabilities	6·94, 4·17	4·71, 2·39, 0·85	21·70, 7·47, 1·73
χ^2-probabilities (approx.)	6·3, 2·8	4·5, 2·3, 0·9	20·5, 5·0, 2·0

The agreement is clearly good enough, considering the very small values of m and n.

6. In the non-random case, if there are n objects, there are $n!$ ways of ranking them, which seems to require $n!$ parameters—generally far too large a number to handle, as Kendall, for example, pointed out at the recent symposium (Moran *et al.* 1950).

But if $n = 3$, there are only five independent parameters. Thus, consider a population in which the probabilities of ranking are:

Ranking	123	321	231	213	312	132
Probability	p_1	p_2	p_3	p_4	p_5	p_6

where $\sum_{i=1}^{6} p_i = 1$. Then the moments of the sample value t of the coefficient τ between rankings drawn from two such populations (the second being distinguished by dashes) are given by

$$E(t^{2l+1}) = q_1q_1' + q_2q_2' + q_3q_3' - \frac{2}{3^{2l+1}}(q_1q_2' + q_1'q_2 + q_2q_3' + q_2'q_3 + q_3q_1' + q_3'q_1),$$

$$E(t^{2l}) = r_1r_1' + r_2r_2' + r_3r_3' + \frac{2}{3^{2l}}(r_1r_2' + r_1'r_2 + r_2r_3' + r_2'r_3 + r_3r_1' + r_3'r_1),$$

where $q_1 = p_1 - p_2, \quad q_2 = p_3 - p_4, \quad q_3 = p_5 - p_6,$

and $r_1 = p_1 + p_2, \quad r_2 = p_3 + p_4, \quad r_3 = p_5 + p_6.$

Clearly, the distribution of an average of a number of such correlations would rapidly tend to normality with a variance that could easily be calculated. Exactly the same result holds, of course, for Spearman's ρ, if one substitutes powers of $\frac{1}{2}$ for $\frac{1}{3}$ in the expressions for the moments. Whilst in principle it is possible to obtain similar results for $n > 3$, no reasonably simple way of writing the expressions is apparent; in the case of four objects nearly fifty parameters would have to be estimated and for $n = 8$, say, almost one hundred thousand are involved. Again, for more than two rankings ($m > 2$), the distribution of a coefficient such as u does not seem amenable to simple treatment even when the rankings are all drawn from the same population and there are three objects only.

Considerable simplification in correlating two rankings is possible if one of them is given *a priori*, i.e. is fixed. If the fixed ranking is put in the 'natural' order 1, 2, 3, etc., the results of the last paragraph, for example, would reduce to

$$E(t^{2l+1}) = q_1 - \frac{q_2+q_3}{3^{2l+1}},$$

$$E(t^{2l}) = r_1 + \frac{r_2+r_3}{3^{2l}} = r_1 + \frac{(1-r_1)}{3^{2l}}.$$

Generally, with one ranking fixed, τ does not distinguish between certain different cases in the other ranking, i.e. all those for which the minimum number of interchanges of pairs of objects which transform them to the given ranking (cf. Kendall, 1948) are equal. For three objects, for example, both rankings 132 and 213 need at least one interchange, the rankings 231 and 312 need two, and the ranking 321 needs three interchanges to transform them to the order 123. All this results in a radical reduction in the number of parameters required, from $n!$ to $N = {}_nC_2$ independent parameters. Thus, in correlating with a fixed ranking, let P_s be the probability of sampling a ranking requiring a minimum of s interchanges to reduce it to the given ranking ($s = 0, 1, 2, \ldots, N$). Then the moments of t are

$$E(t^l) = \sum_{s=0}^{N} \left(1 - \frac{2s}{N}\right)^l P_s,$$

which even for $n = 8$ objects, say, has less than thirty terms, each very simple to estimate if sufficient data are available, as in the problem of m rankings. Here the average of the m coefficients τ with the fixed ranking would rapidly tend to normality, with variance $2(2n+5)/9mn(n-1)$ on the null-hypothesis of random rankings, and otherwise with variance obtainable from the uncorrected moments just given, the probabilities P_s being estimated from the m rankings.

7. One other kind of coefficient may be mentioned in conclusion. Instead of forming

$$u = \sum_{a<b}^{m} t_{ab}/M,$$

where t_{ab} is the rank correlation between rankings a and b, one could consider a coefficient which is insensitive to complete reversals of order in any one ranking, such as

$$u' = \sum_{a<b}^{m} t_{ab}^2/M.$$

The mean and variance under the null hypothesis are easily seen to be given by

$$E(u') = \frac{2(2n+5)}{9mn(n-1)},$$

$$9MN^3 \text{ var. } (u') = \frac{n(n-1)(2n+5)^2(m^2-5m+10)}{144} - \frac{6n^2+21n^2+31n+31}{25}.$$

Apart from testing for the rather peculiar kind of non-random asymmetry implied (cf. example in Kendall, 1948, p. 89), no particular applications of this coefficient are at present apparent. It is mentioned because it seems a rather simpler form of a suggestion made recently by Moran (Moran *et al.* 1950, p. 157).

This note has been written in connexion with work carried out on behalf of the Department of Scientific and Industrial Research (Food Investigation Organization).

REFERENCES

DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, **33**, 129.

FISHER, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc.* A, **121**, 654.

FRIEDMAN, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.* **32**, 675.

KENDALL, M. G. (1948). *Rank Correlation Methods.* London: Griffin and Co.

KENDALL, M. G. & BABINGTON SMITH, B. (1939). The problem of *m* rankings. *Ann. Math. Statist.* **10**, 275.

KENDALL, M. G. & BABINGTON SMITH, B. (1940). On the method of paired comparisons. *Biometrika*, **31**, 324.

MORAN, P. A. P. (1951). *Rank Correlation Methods* (Review). *J.R. Statist. Soc.* A, **114**, 99.

MORAN, P. A. P., WHITFIELD, J. W. & DANIELS, H. E. (1950). Symposium on ranking methods. *J.R. Statist. Soc.* B, **12**, 153.

STUART, A. (1951). An application of the distribution of the ranking concordance coefficient. *Biometrika*, **38**, 33.

LEAST-SQUARES ESTIMATION OF LOCATION AND SCALE PARAMETERS USING ORDER STATISTICS

By E. H. LLOYD, *Imperial College, London*

1. Introduction

In this paper we are concerned with distributions which depend on location and scale parameters only. For such distributions it will be shown that the parameters may be estimated by applying general least-squares theory to an ordered sample, the resulting estimates being unbiased, linear in the ordered observations, and of minimal variance. Explicit formulae are obtained for the estimates, and for their variances and covariance.

The special case of symmetrical distributions is discussed in some detail, and it is shown that for these the estimates are uncorrelated. As examples, the rectangular and the normal distributions are discussed; in the normal case the 'ordered' estimate of the mean turns out to be the sample mean, whereas in the rectangular case the ordered estimates of mean and range are functions of the extreme observations only, with sampling variances of order n^{-2}.

The 'ordered' estimate of the population mean has a sampling variance which never exceeds that of the sample mean; it is equal to that of the sample mean if and only if the row-totals of the variance matrix of the ordered observations all have the same sum. The 'ordered' estimate then becomes the sample mean itself. When the variance matrix does not satisfy this condition the 'ordered' estimate of the population mean has strictly smaller sampling variance than has the sample mean.

2. Existence of least-squares estimates

We consider the estimation of the location and scale parameters μ, σ (not necessarily the mean and standard deviation) of a variate X whose distribution depends on only these two parameters. Let $(X_1, X_2, \ldots, X_n)$ be a sample of n independent observations on X. Arrange the X_r in ascending order of magnitude and denote the ordered set by $(Y_1, Y_2 \ldots, Y_n)$, so that

$$Y_1 \leqq Y_2 \leqq \ldots \leqq Y_n. \tag{2·1}$$

We shall consider unbiased estimates which are linear functions of these ordered observations, and we first prove that the parameters may both be estimated by functions of this type which have minimal variance. To do this we introduce the standardized variates

$$U_r = (X_r - \mu)/\sigma,$$

which may be regarded as independent observations on the standardized variate $U = (X-\mu)/\sigma$, whose distribution is parameter-free. We arrange the U_r in ascending order of magnitude, denoting the ordered set by $(V_1, V_2, \ldots, V_n)$. Then

$$V_r = (Y_r - \mu)/\sigma$$

and

$$V_1 \leqq V_2 \leqq \ldots \leqq V_n.$$

Let
$$\mathscr{E}(V_r) = \alpha_r, \tag{2·2}$$
$$\text{var}\,(V_r) = \omega_{rr}, \quad \text{cov}\,(V_r, V_s) = \omega_{rs};$$

these quantities have known values depending on the form of the parent distribution but not on the parameters μ and σ.

Reverting now to the original ordered observations we clearly have

$$\mathscr{E}(Y_r) = \mu + \sigma\alpha_r, \quad \text{var}\,(Y_r) = \sigma^2\omega_{rr}, \quad \text{cov}\,(Y_r, Y_s) = \sigma^2\omega_{rs}. \tag{2·3}$$

Since the ordered observations have expectations which are linear functions of the parameters μ and σ, with known coefficients, and variances and covariances which are known up to a scale factor σ^2, the least-squares theorem of Gauss and Markoff applies (see, for example, Aitken, 1935). The parameters are therefore estimable by unbiased linear functions of the Y_r, of minimal variance.

[Godwin (1949*a*) has recently proved a similar theorem for the estimation of dispersion alone.]

3. The least-squares estimates of μ and σ

We write the equations (2·3) in matrix form, as follows:

$$\mathscr{E}(\mathbf{Y}) = \mu\mathbf{1} + \sigma\boldsymbol{\alpha},$$

where $\mathbf{Y}$ is the vector of the Y_r, $\boldsymbol{\alpha}$ the vector of the α_r, and $\mathbf{1}$ a vector with unit elements. This equation may be written more compactly as

$$\mathscr{E}(\mathbf{Y}) = \mathbf{p}\boldsymbol{\theta}, \tag{3·1}$$

where $\mathbf{p}$ is the $(n \times 2)$ matrix $(\mathbf{1}, \boldsymbol{\alpha})$, and $\boldsymbol{\theta}' = (\mu, \sigma)$. The variance matrix of the Y_r, i.e. the matrix of variances and covariances, is

$$\mathscr{V}(\mathbf{Y}) = \sigma^2\boldsymbol{\omega}, \tag{3·2}$$

where $\boldsymbol{\omega}$ is the $(n \times n)$ symmetric positive-definite matrix of the ω_{rs}.

The required estimator of the vector $\boldsymbol{\theta}$ of parameters is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{p}'\boldsymbol{\Omega}\mathbf{p})^{-1}\mathbf{p}'\boldsymbol{\Omega}\mathbf{Y}, \tag{3·3}$$

where $\boldsymbol{\Omega} = \boldsymbol{\omega}^{-1}$. The variance matrix of the estimates is $(\mathbf{p}'\boldsymbol{\Omega}\mathbf{p})^{-1}\sigma^2$, where

$$\mathbf{p}'\boldsymbol{\Omega}\mathbf{p} = \begin{pmatrix} \mathbf{1}'\boldsymbol{\Omega}\mathbf{1}, & \mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha} \\ \mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha}, & \boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha} \end{pmatrix}, \tag{3·4}$$

the elements of this matrix being, of course, scalars. The inverse of this matrix is

$$(\mathbf{p}'\boldsymbol{\Omega}\mathbf{p})^{-1} = \frac{1}{\Delta}\begin{pmatrix} \boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}, & -\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha} \\ -\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha}, & \mathbf{1}'\boldsymbol{\Omega}\mathbf{1} \end{pmatrix}, \tag{3·5}$$

where Δ is the determinant of the matrix $\mathbf{p}'\boldsymbol{\Omega}\mathbf{p}$.

Using these results in (3·3) we find for the estimates

$$\hat{\mu} = -\boldsymbol{\alpha}'\boldsymbol{\Gamma}\mathbf{Y}, \quad \hat{\sigma} = \mathbf{1}'\boldsymbol{\Gamma}\mathbf{Y},$$

where $\boldsymbol{\Gamma}$ is the skew-symmetric matrix defined by

$$\boldsymbol{\Gamma} = \boldsymbol{\Omega}(\mathbf{1}\boldsymbol{\alpha}' - \boldsymbol{\alpha}\mathbf{1}')\boldsymbol{\Omega}/\Delta.$$

The variances and the covariance of these estimates are given by

$$\text{var}\,(\hat{\mu}) = \boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}\sigma^2/\Delta, \quad \text{var}\,(\hat{\sigma}) = \mathbf{1}'\boldsymbol{\Omega}\mathbf{1}\sigma^2/\Delta, \quad \text{cov}\,(\hat{\mu}, \hat{\sigma}) = -\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha}\sigma^2/\Delta. \tag{3·6}$$

4. Symmetry properties of ordered observations obtained from a symmetric parent

From now on we restrict our attention to symmetric distributions. For these we shall take the location parameter μ to be the population mean.

The ordered values V_r of observations on the standardized variate U satisfy the inequalities

$$V_1 \leqq V_2 \leqq \ldots \leqq V_n,$$

and therefore also

$$-V_n \leqq -V_{n-1} \leqq \ldots \leqq -V_1.$$

Hence the set $(-V_n, \ldots, -V_1)$ may be regarded as ordered observations on the variate $-U$. Now, since U is symmetrically distributed about a zero mean, its distribution coincides with that of $-U$. Thus the joint distribution of the ordered observations:

$$(V_1, V_2, \ldots, V_n) \text{ is the same as that of the set } (-V_n, -V_{n-1}, \ldots, -V_1); \tag{4·1}$$

in particular, the two sets have the same means and the same variance matrix.

We can express this succinctly in matrix form by introducing the $(n \times n)$ permutation matrix $\mathbf{J}$,

$$\mathbf{J} = \begin{pmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{pmatrix}.$$

It may be noted that $\mathbf{J}$ is symmetric and orthogonal, and its row totals are all unity:

$$\mathbf{J} = \mathbf{J}' = \mathbf{J}^{-1}, \quad \mathbf{J}'\mathbf{1} = \mathbf{1}. \tag{4·2}$$

When it is used as a pre-multiplier, $\mathbf{J}$ has the effect of reversing the order of the rows of the matrix on which it operates. Thus

$$\begin{pmatrix} -V_n \\ -V_{n-1} \\ \vdots \\ -V_1 \end{pmatrix} = -\mathbf{J} \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{pmatrix}.$$

The assertion (4·1) thus means that the vector variates $\mathbf{V}$ and $-\mathbf{JV}$ have the same distribution. In particular

$$\mathscr{E}(\mathbf{V}) = \boldsymbol{\alpha} = \mathscr{E}(-\mathbf{JV}),$$

$$\mathscr{V}(\mathbf{V}) = \boldsymbol{\omega} = \mathscr{V}(-\mathbf{JV}).$$

It follows that

$$\boldsymbol{\alpha} = -\mathbf{J}\boldsymbol{\alpha}, \quad \boldsymbol{\omega} = \mathbf{J}\boldsymbol{\omega}\mathbf{J}.$$

Inverting both members of the last equation we obtain

$$\boldsymbol{\Omega} = \mathbf{J}\boldsymbol{\Omega}\mathbf{J}.$$

5. Estimation of μ and σ for a symmetric distribution

To proceed with the actual estimation we first prove that in the case of a symmetric parent the matrix $\mathbf{p}'\boldsymbol{\Omega}\mathbf{p}$ of (3·4) is diagonal. Its off-diagonal elements are each proportional to $\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha}$, and, on applying the symmetry properties developed above, we have

$$\begin{aligned} \mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha} &= \mathbf{1}'(\mathbf{J}\boldsymbol{\Omega}\mathbf{J})(-\mathbf{J}\boldsymbol{\alpha}) \\ &= -\mathbf{1}'\mathbf{J}\boldsymbol{\Omega}\mathbf{J}^2\boldsymbol{\alpha} = -\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha}. \end{aligned}$$

since $\mathbf{1'J = 1'}$ and $\mathbf{J^2 = I}$, the unit matrix. The element $\mathbf{1'\Omega\alpha}$ is therefore equal to its own negative, and so is zero. The matrix $\mathbf{p'\Omega p}$ is thus diagonal, as is its inverse. Since this inverse is proportional to the variance matrix of our estimates it follows that, in the case of a symmetric parent, $\hat{\mu}$ and $\hat{\sigma}$ are uncorrelated.

The inverse of $\mathbf{p'\Omega p}$ now takes the simple form

$$\text{diag}\,(1/\mathbf{1'\Omega 1}, 1/\boldsymbol{\alpha}'\boldsymbol{\Omega\alpha}),$$

and the estimates become

$$\left.\begin{aligned} \hat{\mu} &= \frac{\mathbf{1'\Omega Y}}{\mathbf{1'\Omega 1}} = \frac{\Sigma\Omega_{ij}Y_j}{\Sigma\Omega_{ij}}, \\ \hat{\sigma} &= \frac{\boldsymbol{\alpha}'\boldsymbol{\Omega Y}}{\boldsymbol{\alpha}'\boldsymbol{\Omega\alpha}} = \frac{\Sigma\Omega_{ij}\alpha_i Y_j}{\Sigma\Omega_{ij}\alpha_i\alpha_j}. \end{aligned}\right\} \tag{5·1}$$

Their variances are

$$\left.\begin{aligned} \text{var}\,(\hat{\mu}) &= \frac{\sigma^2}{\mathbf{1'\Omega 1}} = \frac{\sigma^2}{\Sigma\Omega_{ij}}, \\ \text{var}\,(\hat{\sigma}) &= \frac{\sigma^2}{\boldsymbol{\alpha}'\boldsymbol{\Omega\alpha}} = \frac{\sigma^2}{\Sigma\Omega_{ij}\alpha_i\alpha_j}, \end{aligned}\right\} \tag{5·2}$$

and their covariance is zero. In these formulae the 'Σ' denotes summation with respect to both suffices.

It may be noted that the expected values of *unordered* observations are independent of the scale parameter. If we use unordered observations, therefore, the method of least squares is incapable of providing an optimum linear estimate of the scale parameter; it is certainly possible to find an unbiased quadratic estimate of the population variance, but this will not necessarily have any optimum properties.

6. An example: the rectangular distribution

A simple illustration which is, nevertheless, of considerable interest is provided by the well-known case of the rectangular distribution. We take the parameters μ and σ to be the mean and the standard deviation of the population.

The expectation vector and the variance matrix of the ordered observations can be evaluated explicitly, and we easily find, for a sample of n,

$$\alpha_r = (\sqrt{3})\,(2r-n-1)/(n+1),$$

and

$$\omega_{rs} = 12r(n-s+1)/(n+1)^2(n+2) \quad (r \leqq s).$$

The inverse $\boldsymbol{\Omega}$ of $\boldsymbol{\omega}$ may be shown by induction to be given by

$$\frac{12\boldsymbol{\Omega}}{(n+1)(n+2)} = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & -1 & 2 \end{pmatrix}$$

We have

$$\mathbf{1'\Omega} = \frac{(n+1)(n+2)}{12}(1, 0, 0, \dots, 0, 1),$$

and

$$\boldsymbol{\alpha}'\boldsymbol{\Omega} = \frac{(n+1)(n+2)}{4\sqrt{3}}(-1, 0, 0, \dots, 0, 1).$$

whence $$\hat{\mu} = \tfrac{1}{2}(Y_1 + Y_n),$$

$$\hat{\sigma} = \frac{1}{2\sqrt{3}} \frac{(n+1)}{(n-1)} (Y_n - Y_1).$$

The sampling variances of these estimates are

$$\text{var}(\hat{\mu}) = \frac{6\sigma^2}{(n+1)(n+2)}, \quad \text{var}(\hat{\sigma}) = \frac{2\sigma^2}{(n-1)(n+2)};$$

the covariance is, of course, zero.

Our estimates thus have sampling variances of order n^{-2}, which is in striking contrast to the results obtained from moment estimates calculated from the unordered observations. For example, if we estimate μ by means of the sample mean m, and σ by means of the sample standard deviation s, where

$$ns^2 = \Sigma(X_i - m)^2,$$

we have sampling variances of order n^{-1} only:

$$\text{var}(m) = \sigma^2/n,$$

$$\text{var}(s) = \sigma^2/5n + O(n^{-2}).$$

These variances are larger than those of the 'ordered' estimates by a whole order of magnitude.

Unfortunately, one cannot compare the sampling variances of the 'ordered' estimates with the values given by the Cramer-Rao inequality, since this does not apply to distributions (such as the rectangular) which have discontinuity points at positions determined by the parameters.

It might be noted that attention was drawn to the low asymptotic value $6/n$ of $\text{var}(\hat{\mu})$ by R. A. Fisher, as long ago as 1921 (Fisher, 1921). A few years later Neyman and Pearson (1928) gave a detailed discussion of the sample centre and range.

To conclude the discussion of this example, we might perhaps think it more natural in the case of the rectangular distribution to estimate the extremities λ_1 and λ_2 rather than the mean and the standard deviation. These may be directly estimated by a similar procedure to the above, or, alternatively, we may use the fact that their least-squares estimates $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are connected with $\hat{\mu}$ and $\hat{\sigma}$ by the same linear relationships which connect the corresponding parameters. Since

$$\lambda_1 = \mu - \sigma\sqrt{3}, \quad \lambda_2 = \mu + \sigma\sqrt{3},$$

the estimates are

$$\hat{\lambda}_1 = Y_1 - (Y_n - Y_1)/(n-1),$$

$$\hat{\lambda}_2 = Y_n + (Y_n - Y_1)/(n-1),$$

with $$\text{var}(\hat{\lambda}_1) = \text{var}(\hat{\lambda}_2) = n(\lambda_2 - \lambda_1)^2/(n^2-1)(n+2),$$

and $$\text{cov}(\hat{\lambda}_1, \hat{\lambda}_2) = -(\lambda_2 - \lambda_1)^2/(n^2-1)(n+2).$$

The corresponding unbiased estimate of the range is

$$\hat{\lambda}_2 - \hat{\lambda}_1 = (Y_n - Y_1)(n+1)/(n-1),$$

with sampling variance $$2(\lambda_2 - \lambda_1)^2/(n+1)(n+2).$$

Finally, we remark that $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are jointly sufficient estimators of λ_1 and λ_2.

7. CONDITIONS UNDER WHICH $\hat{\mu}$ HAS A SMALLER VARIANCE THAN THE SAMPLE MEAN

Since the sum ΣY_i of the ordered observations is the same as the sum ΣX_i of the unordered observations, the sample mean may be defined equally as $\Sigma X_i/n$ and as $\Sigma Y_i/n$. It is therefore an unbiased linear compound of the ordered observations. Our estimate $\hat{\mu}$ has, by construction, minimal variance in the class of such estimates, so that its variance is at most equal to that of the sample mean. Hence

$$\text{var}(\hat{\mu}) \leqq \sigma^2/n. \tag{7·1}$$

It is of interest to investigate the conditions under which this relation is a strict inequality. That such a situation can in fact exist is demonstrated by the example of the rectangular distribution.

We first note that the variance matrix $\mathbf{M}$ of any non-degenerate distribution is symmetric and positive-definite. The symmetry property is evident; if $Z_1, Z_2, \ldots, Z_n$ are the variates involved,

$$M_{ij} = \text{cov}(Z_i, Z_j) = \text{cov}(Z_j, Z_i) = M_{ji}.$$

To demonstrate the positive-definiteness consider the variate

$$\lambda_1 Z_1 + \lambda_2 Z_2 + \ldots + \lambda_n Z_n.$$

The variance σ^2 of this expression is a quadratic form in the λ_i, with matrix $\mathbf{M}$. This quadratic form σ^2, being a variance, is necessarily positive for all non-zero $\boldsymbol{\lambda}$. Hence $\mathbf{M}$ is positive-definite.

The variance matrix $\boldsymbol{\omega}$ of our ordered observations $\mathbf{V}$ is thus symmetric and positive definite. It may therefore be written in the form

$$\boldsymbol{\omega} = \mathbf{t}\mathbf{t}', \tag{7·2}$$

where $\mathbf{t}$ is a lower triangular matrix, i.e. $t_{ij} = 0$ for $i < j$. (This is the Choleski resolution. See, for example, Fox, Huskey & Wilkinson (1948).)

Consider now the sum ΣU_i of the unordered standardized observations; its variance is clearly equal to n. Noting that

$$\begin{aligned} \Sigma U_i &= \Sigma V_i \\ &= \mathbf{1}'\mathbf{V} \end{aligned}$$

in matrix notation, we see that

$$\begin{aligned} n = \text{var}(\mathbf{1}'\mathbf{V}) &= \mathbf{1}'\boldsymbol{\omega}\mathbf{1} \\ &= \mathbf{1}'\mathbf{t}\mathbf{t}'\mathbf{1} = \mathbf{h}'\mathbf{h} \\ &= \Sigma h_i^2, \end{aligned} \tag{7·3}$$

say, where $\mathbf{h} = \mathbf{t}'\mathbf{1}$.

Similarly, we find

$$\begin{aligned} \mathbf{1}'\boldsymbol{\Omega}\mathbf{1} &= \mathbf{1}'(\mathbf{t}^{-1})'\,\mathbf{t}^{-1}\mathbf{1} = \mathbf{k}'\mathbf{k} \\ &= \Sigma k_i^2, \end{aligned} \tag{7·4}$$

say, where $\mathbf{k} = \mathbf{t}^{-1}\mathbf{1}$.

Further,

$$\begin{aligned} \Sigma h_i k_i = \mathbf{h}'\mathbf{k} &= \mathbf{1}'\mathbf{t}\mathbf{t}^{-1}\,\mathbf{1} = \mathbf{1}'\mathbf{1} \\ &= n. \end{aligned} \tag{7·5}$$

Now by Schwartz's inequality we have

$$\Sigma k_i^2 \geqq (\Sigma h_i k_i)^2/\Sigma h_i^2, \tag{7·6}$$

whence, using (7·3) and (7·5), $\Sigma k_i^2 \geqq n$,

or $\mathbf{1'\Omega 1} \geqq n$.

But since $\text{var}(\hat{\mu}) = \sigma^2/\mathbf{1'\Omega 1}$, this inequality reduces to

$$\text{var}(\hat{\mu}) \leqq \sigma^2/n,$$

which is, of course, (7·1) again. We are now, however, in a position to state conditions under which this relation becomes an equation. By the well-known properties of Schwartz's inequality, the equality sign in (7·6) applies if and only if the k_i are proportional to the h_i, say

$$k_i = bh_i,$$

for some constant b. When this is so, it follows from (7·4) and (7·5) that $b = 1$.

Thus (7·6) is an equation if and only if

$$\mathbf{k} = \mathbf{h},$$

or $$\mathbf{t'1} = \mathbf{t}^{-1}\mathbf{1},$$

or, since $\mathbf{tt'} = \boldsymbol{\omega}$, $$\boldsymbol{\omega 1} = \mathbf{1}. \tag{7·7}$$

The interpretation of this condition is that, in each row (and, by symmetry, each column) of the matrix $\boldsymbol{\omega}$, the sum of the elements is unity.

When the condition holds, not only is the variance of $\hat{\mu}$ equal to that of the sample mean, the estimate $\hat{\mu}$ in fact becomes the sample mean. For (7·7) implies that

$$\boldsymbol{\Omega 1} = \mathbf{1},$$

so that
$$\hat{\mu} = \frac{\mathbf{1'\Omega Y}}{\mathbf{1'\Omega 1}} = \frac{\mathbf{1'Y}}{\mathbf{1'1}}$$
$$= \frac{\Sigma Y_i}{n} = \frac{\Sigma X_i}{n}.$$

When the condition is not satisfied,

$$\text{var}(\hat{\mu}) < \sigma^2/n.$$

Thus, when the row totals of the variance matrix of the ordered observations are all equal, the 'ordered' estimate $\hat{\mu}$ coincides with the sample mean; when, however, the row totals of the variance matrix are not all equal, $\hat{\mu}$ is a better estimate that the sample mean ('better' in the sense of having smaller sampling variance).

An example of the latter case has already been discussed. An example of the former case is provided by the normal distribution.

8. The normal distribution

It is known from general theory that no estimator of the mean of the normal distribution can be better (in the above sense) than the sample mean. The ordered estimate must therefore coincide with the sample mean for this distribution.

(This, incidentally, proves that, for the standardized normal distribution, every row of the variance matrix of the ordered observations adds up to unity.)

The 'ordered' estimate of the standard deviation, however, cannot be dealt with so easily; the expected values, variances and covariances of the ordered observations have to be computed before the formulae (5·1, 5·2) can be used. Godwin (1949*b*) has tabulated the required figures for sample sizes up to $n = 10$, and, in another paper (1949*a*), has compared the efficiencies of the 'ordered' estimate with various other estimates of σ, the efficiency of the ordered estimate, for $n \leqslant 10$, being at least 98·83 %.

I am grateful to Prof. M. G. Kendall for suggesting a number of improvements to the original draft of this paper.

REFERENCES

AITKEN, A. C. (1935). On least squares and linear combination of observations. *Proc. Roy. Soc. Edinb.* **55**, 42.

FISHER, R. A. (1921). On the mathematical foundations of theoretical statistics. *Philos. Trans.* A, **222**, 309.

FOX, L., HUSKEY, H. D. & WILKINSON, J. H. (1948). Notes on the solution of algebraic linear simultaneous equations. *Quart. J. Mech. Appl. Math.* **1**, 149 (esp. pp. 159 ff.).

GODWIN, H. J. (1949*a*). On the estimation of dispersion by linear systematic statistics. *Biometrika*, **36**, 92.

GODWIN, H. J. (1949*b*). Some low moments of order statistics. *Ann. Math. Statist.* **20**, 279.

NEYMAN, J. & PEARSON, E. S. (1928). On the use and interpretation of certain test criteria. *Biometrika*, **20**A, 175.

REGRESSION, STRUCTURE AND FUNCTIONAL RELATIONSHIP. PART II

By M. G. KENDALL

Division of Research Techniques, London School of Economics

1. In the first part of this paper I reviewed the theory of regression and the determination of functional relationship when both variables are subject to error. In this continuation I examine some of the problems arising in the structural analysis, but before doing so add a few further comments on the problem of determining functional relationship.

Further remarks on functional relationship

2. In 1940 Wald propounded a new solution of the problem of fitting a straight line to data when both variables are subject to error. He obtained estimators and confidence intervals for the parameters α_0 and α_1 in the equation

$$Y = \alpha_0 + \alpha_1 X, \tag{1}$$

where, as in Part I, the variables Y and X are subject to errors of observation ϵ and η. Wald made the usual assumptions about the independence of ϵ and η and of successive values of each. The essence of the method is to divide the observations (assumed even in number, say, $2m$) into two groups of m and to construct as an estimator of α_1 the statistic

$$a_1 = \frac{\sum_{j=1}^{m} y_j - \sum_{j=m+1}^{2m} y_j}{\sum_{j=1}^{m} x_j - \sum_{j=m+1}^{2m} x_j}, \tag{2}$$

x and y being the observed values of X and Y. He also requires a new type of assumption; the limit inferior of

$$\frac{1}{2m}\left| \sum_{j=1}^{m} E(x_j) - \sum_{j=m+1}^{2m} E(x_j) \right| \quad (m \to \infty) \tag{3}$$

is positive. Under these conditions the estimator (2) is consistent as m increases. It appears at first sight as if Wald's solution contradicts the statement of §42 (Part I) to the effect that no solution is possible without some assumption concerning the relative magnitudes of the errors. But condition (3) is really equivalent to some assumption of this kind. We have, in the general case, no reason to suppose that it is satisfied, for $x_1 \dots x_m$ can be any of the $2m$ values of x observed. If we take them to be the smallest values they are not independent of the errors of observation; and if we take them as the values with the smallest expectations we require to know those expectations, or at least to know that the expectations obey (3). To decide this point by inspection of the data is not always possible; and Wald himself notes (pp. 297–8) a remark by Hotelling that if ϵ, η are normal and the values X are themselves chosen randomly from a normal population, the expression (3) converges in probability to zero for any set of x's which is defined independently of the error.*

* Geary (1942) observes in a footnote on p. 69 that Wald's method is applicable when the 'variables' are distributed on the normal surface of error, but I think he has overlooked this remark; and, indeed, the implications of the condition (3) are very far from obvious.

3. Nevertheless, Wald's paper is an important contribution. He showed that whereas in earlier treatments a knowledge of the parent parameters $\text{var}\,\epsilon$ and $\text{var}\,\eta$ is required—or at least a knowledge of their ratio—a much more general assumption of the type associated with expression (3) will suffice to provide consistent estimators in cases where we can objectively split the values X into two groups obeying it. From a somewhat similar point Theil (1950) has recently developed a theory of distribution-free regression. Speaking somewhat loosely, one might say that the problem is resolved if the errors in the x-variable are not large enough to disturb the orders of the variates $\sum_1^m x_j$ and $\sum_{m+1}^{2m} x_j$, i.e. that the first is or is not greater than the second according as the expectation $\sum_1^m E(x_j)$ is or is not greater than $\sum_{m+1}^{2m} E(x_j)$. In a more refined model Theil requires the assumption that

$$E(x_1), E(x_2), \ldots, E(x_{2m})$$

are in the same order (or the same order inverted) as $x_1 \ldots x_{2m}$, which is always so if the observational errors are less than the intervals between the expectations.

The problem of fitting functional relations is thus soluble under certain assumptions concerning the order of the values of the underlying variables. Such assumptions are often plausible or can be made so by spacing out the observations; but they have to be made nevertheless.

4. Berkson (1950) has recently claimed to be able to derive an unbiased estimator of the constants in a regression relationship where the variates are both subject to error. It would appear as if his results (pp. 170–3) are equivalent to equations (42) and (46) of Part I of this paper. But it seems to me that he applies them in cases where I should regard their uses as unwarranted, namely, in the estimation of constants in linear relationships. In the notation of §40, Part I, suppose we have two observables, x', y', given by

$$\left.\begin{aligned} x'_X &= X + v_X, \\ y'_Y &= Y + w_Y. \end{aligned}\right\} \tag{4}$$

If we estimate α_1 in equation (1) by least squares applied to x' and y' we find

$$a_1 = \Sigma y'x'/\Sigma x'^2. \tag{5}$$

Berkson now regards x' as 'preassigned' or, as I should say, fixed, and argues in this way (p. 173): we have, from (4),

$$\begin{aligned} y' &= Y + w \\ &= \alpha_0 + \alpha_1 X + w \\ &= \alpha_0 + \alpha_1 x' - \alpha_1 v + w. \end{aligned}$$

Hence $$E(a_1) = E[\{\Sigma\alpha_0 x' + \alpha_1 \Sigma x'^2 - \alpha_1 \Sigma x'v + \Sigma wx'\}/\Sigma x'^2].$$

For fixed x'

$$\left.\begin{aligned} E\Sigma x' &= 0 \quad \text{(by convention)}, \\ E\Sigma x'v &= 0 \quad \text{(for all } X\text{)}, \\ E\Sigma x'w &= 0 \quad \text{(for all } X, Y\text{)}, \end{aligned}\right\} \tag{6}$$

and hence, from the expression for $E(a_1)$,

$$E(a_1) = \alpha_1. \tag{7}$$

5. Apparently Berkson regards this result as valid in estimating α_1 in an experiment where the x' are assigned values. By assigning values to a random variate he seems to mean that if we fix x' we do not fix X, but commit an error v_X in attempting to do so. Thus, in an experiment on dosage mortality, if we weigh out what purports to be 2 g. of a drug, we may commit an error; but if we nevertheless carry out an estimation by equation (5) as if no error were committed, we shall still get an unbiased estimation, the true value (actual weight X) being unknown and the error v_X being unknown, but their sum *ex hypothesi* being 2 g., the observed x'. All that is required is that equations (6) shall be satisfied.

But this requirement seems to me to be incapable of being satisfied. If we fix x'_X and still require v_X to be a variate then X must be a variate, which is impossible. If x'_X is not fixed then from (4)

$$E(x'_X v_X) = E(X v_X + v_X^2),$$

which cannot vanish if $E(v_X) = 0$.

6. Berkson, as it seems to me, inadvertently confuses the issue by entitling his paper 'Are there two regressions?' (Frisch does the same by speaking of a 'true regression' instead of a functional relationship.) The answer, in my terminology, is that there are certainly two regressions. If we are estimating constants in a linear relationship (which is nothing to do with regression analysis in the classical sense) and use a least-squares technique, then we cannot obtain an unbiased estimator, when there is experimental error in the predicated variable, unless we introduce auxiliary assumptions. The uniqueness of the linear relationship is not in question.*

7. Perhaps it may remove yet another source of confusion if I refer at this point to some work by Geary (1942, 1943). Geary is concerned with inherent relations between *variates*. If we have a number of variates x subject to errors (of observation, say) ϵ, and if the ϵ's are independent of the x's and of each other then the product-cumulants of the set $x'(= x+\epsilon)$ are the same as those of the x's. If the x's are related by equations such as

$$\sum_{j=1}^{k} \alpha_j x_j = \text{a constant}, \tag{8}$$

then it is possible to write down linear equations in the α's and the product-cumulants, which can be calculated from the observed x'. This gives estimates of the α's in terms of observable quantities without making any demands on the nature of the error except the usual conditions of independence. Unfortunately, the method breaks down, as Geary observes, when the x's are jointly normal because the cumulants all vanish for order greater than 2, or if they are independent, for then the product-cumulants all vanish. It is to be noted also that if x' is normal and x, ϵ are independent, then each must also be normal. The method applies, then, only where the errors or the x's are not normally distributed. It is true that most distributions are not exactly normal, but if they are close to normality a very large sample would be required to detect any deviation. But, in any case, Geary's method applies only to cases of relationships between variates. Where functional relations are concerned there are no cumulants of the unobservable variables.

Geary's results, in fact, relate to structural equations, which I now consider.

Structural relations

8. In studying the joint variation of a complex of variables or variates we often wish to consider the case when some of them are connected by functional relationships, particularly

* See the note at the end of this paper.

of the linear type (since these are simpler to handle). We may have functional relations between variables of the kind represented in equation (1), or functional relations between variates such as

$$y = \beta_0 + \beta_1 x. \tag{9}$$

These relations are usually suggested by some prior theoretical analysis. I have noted that we cannot have mixed equations such as

$$y = \beta_0 + \beta_1 X, \tag{10}$$

although we may have, for example,

$$y_X = \beta_0 + \beta_1 X + \epsilon_X, \tag{11}$$

which means that the variate y, distributed in a form which is dependent on some variable X, is linearly related to another variate ϵ which may or may not depend on X.

9. In writing down these relations we may follow two quite different courses. The point may be illustrated by a simple economic example in which we postulate that quantity sold (Q) is linearly related to price (P) by a relation

$$Q = \alpha_0 + \alpha_1 P. \tag{12}$$

This is an equation suggested, as a rough approximation, by economic analysis, which further suggests, and indeed almost requires, that the sign of α_1 should be negative. Suppose that we wish to compare this equation with the results of a set of observations on the quantity of a commodity sold at various prices. We do not expect that for each pair of observations equation (12) will be exactly obeyed. Wherein lies the source of the error? The first possible procedure, called by Geary (1948) the 'error-in-variable' approach, is to suppose that our observations on Q and P are affected by errors of observation so that instead of given values of Q and P we observe, say, values of $Q + \epsilon_Q$ and $P + \eta_P$. In short, we have the case of a functional relationship with observations subject to error. The second procedure is to regard equation (12) as a condensed version of an equation from which a number of variables (such as income, prices of substitute commodities and so forth) have been omitted, so that the true equation should be the functional relation

$$Q = \alpha_0 + \alpha_1 P + f(U_1, U_2, \ldots). \tag{13}$$

We further suppose that the function f behaves like a stochastic error term and hence wish to write something representing an 'error-in-equation' approach such as

$$Q = \alpha_0 + \alpha_1 P + \epsilon. \tag{14}$$

Unfortunately, this is a meaningless expression, but we can transform it into something possessing a meaning by writing

$$q = \beta_0 + \beta_1 p + \epsilon \tag{15}$$

or

$$q_P = \beta_0 + \beta_1 P + \epsilon_P. \tag{16}$$

In the first of these we have completely altered the model. The function $f(U_1, U_2, \ldots)$ can only be represented as a random variate if we postulate something about the class of values which we shall observe in practice. We are therefore imposing a requirement on the behaviour of the events we are to consider. Our quantities and prices are now themselves random variates and we have to imagine them as possessing a frequency distribution. The alternative model, represented by (16), allows the quantity to be a random variable for any fixed price.

We note, however, that (16) is an equation such as we should derive as a conditional distribution by fixing P in (15), and if ϵ_P has zero mean

$$E(q_P) = \beta_0 + \beta_1 P, \tag{17}$$

which is the regression of q on p in (15).

10. We may call a set of equations in variables or variates a *structural set*, and describe the analysis of them a *structural analysis*. The distinction between regression and structure is that in the first we are interested in functional relationship between means of dependent variates and predicated variables, as in equation (17), whereas in the second we are interested in relationships between variates, as in equation (9). We may also admit into a structural set equations such as (16) where the variates depend on a variable entering explicitly into the equations. These distinctions may, perhaps, seem somewhat overdrawn, but they are worth making because of the differences in technique which are required to analyse the situations to which they correspond.

11. At this point some further definitions are desirable:

*Endogenous and exogenous elements.** In dynamic economics, and indeed in any quantifiable system under scientific examination, we have certain elements which are inherent in the system and others which represent the influence of factors impinging on the system from without. Prices, quantities produced, and income are inherent in the economic system and are *endogenous*. Rainfall, temperature and warfare are *exogenous*. In a laboratory experiment on Boyle's law pressure and volume are endogenous; temperature, magnetic forces, and earth tremors, are exogenous. The classification is a useful one, but I am doubtful whether it would stand up to logical analysis unless we are prepared to postulate something akin to causality. It is reasonable to suppose that rainfall may affect demand but demand will not affect rainfall, or that earth tremors will affect our instrumental readings, but our instruments will not generate earthquakes. The difficulty is that if the two elements, exogenous and endogenous, are concomitant, no statistical analysis can distinguish which is causal to the other. The distinction between their natures must therefore be made on prior, non-statistical grounds. Fortunately, it does not seem necessary to attempt to make the distinction very precise. For present purposes it is enough to remark that exogenous variables are observable; or so I understand it, for if they were not they would be consolidated into an error term.

Predetermined elements. In some circumstances, at least, then, we may regard the exogenous elements as predetermined like a fixed predicated variate. There is a second class of elements which are essentially predetermined, namely, those endogenous variates which are lagged. In an equation giving production at time t in terms of price at time $t-1$ we may regard price as a known element. If production at time t depended on price at time t and price at time $t-1$, the latter would be predetermined but the price at time t would not, unless the form of relationship were such that the price at t were regarded as predicated.

Latent elements. In the equations determining the behaviour of the system there will appear certain elements which are not observable, or at least are not observed. They appear explicitly in the formulation of the problem, but statistical evidence concerning them is indirect. In the terminology of Koopmans (see, for instance, Koopmans & Reiersöl, 1950) the unobservable variates are said to be *latent*. As I shall use the term, error variates, though unobservable, are not latent. Latency refers only to those variates (like g in factor analysis

* I use the word 'element' in this context to denote indifferently variables or variates.

or demand in economics) which are supposed to measure some real quantifiable element in the system.

12. Thus, in the light of prior analysis or information about the system, we may write down a set of equations which determine its motion. For example, if we take the error-in-equation approach, with variates not subject to experimental error, we have in the notation of Monograph 10 of the Cowles Commission (Koopmans, 1950) the typical equation

$$\sum_{i=1}^{G}\sum_{\tau=0}^{t'}\beta_{i\tau}y_i(t-\tau)+\sum_{k=1}^{K}\sum_{\tau=0}^{t'}\gamma_{k\tau}z_k(t-\tau)=u(t), \tag{18}$$

where the y's are endogenous and the z's are exogenous variates. There are, say, g equations of this kind. $u(t)$ is the 'error' term in the equation which is known as the *disturbance*.

We may write equation (18) more simply. First of all, if D is an operator lagging the time element by unity, i.e. if

$$Df(t)=f(t-1),$$

we may replace $\Sigma\beta_{i\tau}y(t-\tau)$ by $\beta y(t)$, where β is now a polynomial in D. Secondly, using a matrix notation, we may write the set of equations such as (18) in the form

$$By+\Gamma z=u, \tag{19}$$

where B is a $g\times G$ matrix, Γ is a $g\times K$ matrix and y, z, u are column vectors (matrices of orders $G\times 1, K\times 1, g\times 1$, respectively). If $G=g$ (and we need not seriously consider any other case, which would imply either that we have more latent variables than equations or vice versa) the solution for the y's in terms of z's and u's is then simply

$$y=-B^{-1}\Gamma z+B^{-1}u, \tag{20}$$

where B^{-1} means the function of the coefficients β_i together with the operator D obtained by inverting the matrix as if D were an algebraic quantity and then expanding any denominator terms in D as an infinite series of powers of D by binomial processes. We require for the validity of (16) that B_0, the matrix of terms independent of D, shall be non-singular.

13. It is, however, of some importance to remark that the object of considering these equations is not to solve for the latent elements y but to estimate the constants β and γ. To that extent the problem resembles the determination of regression coefficients or constants in a functional relationship. It may fairly be asked why, when faced with a stochastic complex of this type, we do not treat it simply as a multiple regression problem expressing each of the endogenous unlagged elements in terms of predetermined elements. I return to this point in §28.

14. The disturbances $u(t)$ are not observable. It is supposed, in general, that they have a joint probability distribution with finite moments up to at least the second order. In simpler models it is also assumed that they are independent among themselves, that successive values of any one are independent and that they are independent of the K exogenous elements. In any case, of course, it follows that the G endogenous elements are also random variates connected with the u's by g equations of type (18), provided that we remember that any endogenous element with $\tau>0$ is fixed at time t.

15. The general problem of structural analysis may then be stated as follows: on prior grounds we write down equations of type (18) which we regard (at least to a first approximation) as determining the system. We next have to examine whether the system is capable

of solution in the sense that it permits of the estimation of any or all of the parameters β and γ. If it does, we perform the estimation and consider whether the values obtained are reasonable in the light of prior knowledge. If they are, the resulting model fits the data and we can proceed to test it on new data by examining its value in prediction or otherwise. If the data do not fit the model, or if it fails as a predictor, we have to begin again and to revise the system of equations..

16. This leads us to a discussion of 'identifiability', that is to say, to the possibility of uniquely determining the parameters β and γ. Four cases arise:

(*a*) no parameters can be determined. The system is then said to be completely unidentifiable;

(*b*) some but not all parameters can be determined. The system is then partially identifiable and those parameters which can be determined are identifiable;

(*c*) all the parameters can be determined and there are only just enough equations to enable determination to be carried out. The system is then completely identifiable.

(*d*) there are more elements present than are necessary for complete identifiability. The system is then over-identified. (This does not mean that the equations are inconsistent among themselves, but only that we could do without some of the exogenous elements and still have an identifiable system.)

17. Identifiability is not a matter of sample size. Even if we have an indefinitely large sample and for all practical purposes know the distribution functions of the disturbances, it may still arise. In fact, it is not, in full generality, solely a statistical problem; but I shall be concerned with it only in a statistical context.

To take a case so simple as to be almost trivial: suppose we have a population which is composed of the sum of two independent variates y and z, each with known mean but unknown variance. It is clear that, however many observations we take on the sum $y+z$, we can determine only the variance of that sum, say v_1+v_2, and cannot determine v_1 and v_2 separately. These two parameters are then unidentifiable.

Consider now the economic case wherein we have a demand equation expressing demand-price (*d*-*p*) relationship

$$d-\beta_0-\beta_1 p=\epsilon, \tag{21}$$

a supply equation (*s*-*p*) relationship

$$s-\gamma_0-\gamma_1 p=\eta, \tag{22}$$

and, for economic reasons, identify both supply and demand with quantity sold:

$$s=d=q. \tag{23}$$

My personal opinion is that the economist is creating a difficulty for himself by adopting equation (23) in conjunction with (21) and (22), but let us pass that over. At this stage it is enough to observe that the practice is generally accepted and leads to an identification problem. In fact we have

$$\left.\begin{aligned} q-\beta_0-\beta_1 p&=\epsilon,\\ q-\gamma_0-\gamma_1 p&=\eta. \end{aligned}\right\} \tag{24}$$

We postulate nothing about ϵ and η except that they are stochastic variates. We can then clearly form an infinite number of pairs of equations such as

$$\left.\begin{aligned} (1+l)q-(\beta_0+l\gamma_0)-(\beta_1+l\gamma_1)p&=\epsilon+l\eta,\\ (1+m)q-(\beta_0+m\gamma_0)-(\beta_1+m\gamma_1)p&=\epsilon+m\eta, \end{aligned}\right\} \tag{25}$$

and these sets are indistinguishable from one another so far as concerns the observations, unless we observe or postulate something about the disturbances ϵ and η. Thus we cannot estimate the constants and the system is unidentifiable.

We may even have situations where a good deal is postulated about the disturbances without resolving the unidentifiability. For instance, in the case of equation (24), ϵ and η may be known to be normal; but then so will be $\epsilon + l\eta$ and $\epsilon + m\eta$. Or they may be supposed normal and independent; but it will still be possible to have values of l and m (depending on the variances of ϵ and η) such that $\epsilon + l\eta$ and $\epsilon + m\eta$ are normal and independent. To make the system identifiable we have either to postulate that ϵ and η will have such distribution laws that this kind of thing cannot happen or to introduce an exogenous variate into one of the equations at least. I consider these possibilities in more detail below.

18. In the foregoing I have spoken of identifiability in relation to the determination of parameters in linear equations. A more general definition, adapted so as to cover both parametric and non-parametric specifications, is given by Koopmans & Reiersöl (1950) following Hurwicz's paper in Monograph 10 (1950) on a 'generalization of the concept of identification'. We begin with a specification of the system, such as a set of equations of type (18), expressing what we suppose to be the relations between the various elements, and containing certain unknown features, such as the constants β and γ. A *structure* is defined by Koopmans & Reiersöl as this specification together with a specification of a particular probability distribution of the latent variates. A *model* is a set of structures. More usually we think of these two ideas in the reverse order. The general specification is called the model, and the structure is one of the particular values it assumes when the unknown parameters and probability distributions are determined. The order of presentation of the two ideas is not of great importance; but it is very important to observe that the model, in this sense, specifies nothing about the disturbances except that they are stochastic. It is the structure which specifies their distributions, either directly or implicatively by specifying the constants β and γ and the distribution of the y's.

19. In the general parametric case represented by equation (18) an equation is unidentifiable if and only if it is possible to construct another equation, by linear combination of the equations in the model, which also conforms to the specification constituting that equation. This is evidently sufficient. The article by Koopmans, Rubin and Leipnik in Monograph 10 shows that it is also necessary. Only linear combinations can bring about unidentifiability except in some very special cases when there is functional dependence among lagged endogenous variates.

It follows immediately that *when nothing is specified about the disturbances* a necessary and sufficient condition for the identifiability of an equation A in a linear model, which is subject to the condition of specification that certain endogenous elements are absent from A, is that if we form determinants of order $G-1$ from the coefficients with which those excluded variables occur in the other equations of the model, one determinant at least does not vanish. I refer to this below as the rank condition.

For example, in the simple case of equations (24) the parameters are all identifiable if we add an exogenous variate x to one equation, giving, say,

$$\left.\begin{aligned} q - \beta_0 - \beta_1 p &= \epsilon + x, \\ q - \gamma_0 - \gamma_1 p &= \eta. \end{aligned}\right\} \tag{26}$$

If we suppose, without loss of generality, that the disturbances and x have zero mean it is easy to see that consistent estimators of the constants are given by

$$\left.\begin{aligned} b_0 &= \gamma_0 = 0, \\ \Sigma qx - c_1 \Sigma px &= 0, \\ \Sigma qx - b_1 \Sigma px &= \Sigma x^2, \end{aligned}\right\} \tag{27}$$

where summation is over sample values, or, in the limit, denotes moments of the second order. It is to be remembered that x is independent of ϵ and η.

20. It may be useful at this point to exemplify the case of over-identification. If we have two exogenous variates x and y we might have, instead of (25), such a system as

$$\begin{aligned} q - \beta_0 - \beta_1 p &= \epsilon + x + y, \\ q - \gamma_0 - \gamma_1 p &= \eta. \end{aligned}$$

The system would be identifiable if either x or y were absent and hence is over-identified. The point of distinguishing the over-identified case is that it modifies our methods of estimation. In fact, if x and y are independent we have, on multiplying the first equation by x and y respectively and taking expectations,

$$\left.\begin{aligned} E(qx) - b_1 E(px) &= E(x^2), \\ E(qy) - b_1 E(py) &= E(y^2), \end{aligned}\right\} \tag{28}$$

and similarly, on multiplying the second,

$$\left.\begin{aligned} E(qx) - c_1 E(px) &= 0, \\ E(qy) - c_1 E(py) &= 0. \end{aligned}\right\} \tag{29}$$

These equations are not, in general, consistent, and we therefore require some rule of procedure in estimation which avoids the conflict and, so to speak, combines the two estimators of b_1 in (28) or of c_1 in (29) into a best estimator. I revert to the point in §28.

21. Let us return to consider the other possibility of resolving the identifiability problem, namely, the specification of the nature of the disturbances. It has been noted above in §16 that we may impose conditions of normality and independence without necessarily making the system identifiable. The question arises: What are the necessary and sufficient conditions to be placed upon the disturbances for the system to become identifiable? From a consideration of the rank condition (§19) this amounts to asking: What specifications on the disturbances are equally obeyed by linear functions of them?

Normality, as is well known, is not the only distributional property which is reproduced by linear functions. The Cauchy and Poisson laws behave in the same way for certain linear combinations. But the normal distribution is the only continuous one with finite second moments with the required property of stability, and if (as seems inevitable for practical discussions) we limit ourselves to such types of distributions we are confining ourselves effectively to the normal case, provided that the disturbances are independent. Thus we may say that if we postulate that the disturbances are independent and do not follow a stable law, the system is identifiable. Unfortunately, this is not much of a help. We usually do not know what the law of the disturbances is or whether they are independent; even if we assume independence we usually assume normality in order to resolve the estimation problem; and in any case our sample is rarely large enough for use to be able to distinguish

normal from non-normal variation in the disturbances. It would seem that in practice we are forced back on to the alternative method of introducing exogenous variates into the equations.

22. Nevertheless, there remain in this field a number of interesting theoretical problems for solution. There seems to be some essential relation between normality and indeterminacy in linear systems (as exemplified by Geary's results and the unidentifiability results referred to above). Normality implies indeterminacy. The question is how far the converse is true, i.e. that only normality implies indeterminacy. In terms of the identifiability problem, suppose that we require of our disturbances that they follow a particular law but are not necessarily independent; how far will linear functions obey that law (*a*) for a finite set of linear functions and (*b*) for an infinite set of linear functions?

It would be pleasant to be able to prove a general theorem to the effect that if $u_1, u_2, \ldots, u_k$ all have the same law of distribution and $\Sigma\lambda_j u_j$ also has that law for arbitrary λ's, then the u's are normal. Unfortunately, this is not true. Lindley (1950) has recently shown, for example, that if u_1 and u_2 have finite variances and zero means then $\lambda u_1 + u_2$ is distributed in the same form for all λ's if and only if u_1 and u_2 have an elliptical distribution. Broadly speaking, this means among other things that the density function of u_1 and u_2 is constant over ellipses $au_1^2 + bu_2^2 =$ constant. More specifically, it is necessary and sufficient for such a distribution to exist that, if $\phi(t)$ is the characteristic function of u_1 the function

$$G(R) = \lim_{T\to\infty} \int_0^T t\,dt \int_0^R J_0(tR)\,R\,dR\,\phi(t) \tag{30}$$

shall be a distribution function.

23. We have got rather a long way from the problem of identifiability as it originally arose, and I return to a more general discussion. The basic concept involved concerns the extent to which an examination of the data (however extensive) will permit of a distinction between certain classes of hypotheses. However, identifiability in the present sense is not the same thing as distinguishability. Suppose, for example, we find that a series can be represented by a Yule autoregressive equation

$$u_{t+2} + \alpha u_{t+1} + \beta u_t = \epsilon_{t+2}, \tag{31}$$

where the ϵ's are random. Such a series could equally well be generated by

$$u_{t+1} + \rho u_t = \eta_{t+1}, \tag{32}$$

where the η's themselves obey

$$\eta_{t+1} + \tau\eta_t = \epsilon_{t+1}, \tag{33}$$

provided that

$$\left.\begin{aligned} \alpha &= \rho + \tau, \\ \beta &= \rho\tau. \end{aligned}\right\} \tag{34}$$

How, then, do we distinguish between two cases? Not on the basis of observation of the u's. So far as concerns the observed series, the generating scheme may be either of the Yule type with random disturbances or the Markoff type with disturbances which are themselves of the Markoff type. The hypotheses are observationally indistinguishable, and indeed are equivalent unless we depart from the attitude that they are descriptive and that the constants α, β, ρ, τ represent some property of the system such as would, for example, enable us to write down structural equations. In the latter case there is a difference: in (31) there are two endogenous constants, whereas in (32) and (33) one constant is endogenous and

the other exogenous. This would be important if we were interested in the prediction of behaviour when structural parameters alter their value, for the exogenous parameter τ is not in general capable of alteration. The indistinguishability of the two models is not a matter of identifiability. It depends on the prior classification into endogenous and exogenous elements. Identifiability arises in connexion with lack of distinction *within the same model.*

24. One last point on identifiability. It has been pointed out above that the difficulties are not removed by increasing the number of observations. They are, as it were, inherent in the formulation of the model. At first sight it appears contradictory for Koopmans & Reiersöl to say (1950, p. 170) that identifiability is subject to statistical test. Indeed, it would be so if they meant that we can determine from the observations whether a *model* is identifiable, that is to say, whether a structural system is identifiable for arbitrary values of the parameters. What they mean, if I have interpreted them correctly, is that there may be certain parameter values which render a model identifiable and others which leave it unidentifiable, and that it is possible to ascertain by examination of the data whether a given parameter falls into one class or the other. This, I think, amounts to saying, in the parametric case, that if we can estimate the parameters we can also test the significance of the estimates and say whether the vanishing of the determinants required by the rank condition (§18) is possible or not within acceptable sampling limits.

Estimation

25. I proceed to consider the estimation of the parameters in equations of type (18). Of the two methods which have been put forward, one (the maximum-likelihood approach) requires a specification to the effect that the probability distribution of the disturbance is known at least in general form; the other (least squares) must either cast back to the maximum-likelihood method for normal disturbances or adopt the principle of least squares itself as a principle of inference. This is reasonable enough when conditions are such that we can apply Gauss's theorem that least-squares solutions in linear problems give minimal variance to the estimators. But where time-lags are present the theorem may not apply, and if we use it we are making a non-proven extension of the theorem. The general problem of formulating a valid system of estimation when there are autocorrelations in the residuals has not been solved.

26. The customary assumption to make is that the disturbances are distributed in the multivariate normal form, and it is noteworthy that in such circumstances the maximum-likelihood estimators are independent of the dispersion matrix of disturbances, so that the problem becomes determinate without any specific assumption concerning that matrix. The actual process of solution may be troublesome, arithmetically speaking, but does not raise any new theoretical difficulties so far as I am aware.

27. The explicit solution of a complete set of structural equations, however, is frequently unnecessary, in the sense that we are interested in only one of the equations, and inconvenient, in the sense that some of the variates occurring elsewhere may be difficult to measure. We are not justified, in general, in writing down the one equation and ignoring the others; such a procedure may introduce bias (as was emphasized by Haavelmo (1944)). But having written down the complete system we *are* justified in considering whether we can estimate the coefficients in one equation without having to estimate uninteresting coefficients

occurring in other equations. Such a method has, in fact, been devised by T. W. Anderson (see paper IX in Monograph 10). It is known as the 'reduced form method', or, more descriptively but far less conveniently, as 'the limited-information maximum-likelihood method'. Essentially the process is a return to the regression approach. By a solution of type (20) we express the endogenous elements under examination in terms of a regression on the predetermined variables. Reference may be made to Anderson's paper in Monograph 10 or to Anderson & Rubin (1949) for the details.

28. It is very natural to inquire why we do not use equation (20) in finding the solution of the general structural set, that is to say, why we do not treat the problem as one in the determination of simultaneous regression equations. The answer is that the methods are equivalent when the system is just identified by its exogenous variates. The only difference, if I have understood the situation correctly, arises when the system is over-identified, in which case the ordinary maximum-likelihood method is better, but the limited-information method gains in ease at the expense of ignoring relevant information. In the only practical case where such equations have been worked out in great detail, Klein's study (1950) on the U.S.A. economy, the reduced-form solutions, are very close to the maximum-likelihood solutions for the complete set.

29. Clearly much more critical study is necessary on practical data before we can say that the approach through structural equations is fruitful. Conceptually it seems to me sound enough so far as concerns the use of prior knowledge or analysis in model-building, but the error-in-variable approach may prove to be better. Practical results in macro-economics have not yet been very encouraging. It may be that econometricians have tried to do too much in formulating equations concerning a whole economy instead of concentrating on a single commodity or a small local market. Or it may be that the key to the successful use of the methods lies in our treatment of the disturbance functions, and that we cannot sweep everything which is unknown or inconvenient into a disturbance term and then assume that it behaves like an observational error. What are required here, apart from more extensive data, are (*a*) methods of testing the disturbance terms to see how far our hypotheses concerning them are plausible, (*b*) methods of estimation which are partly independent of the nature of the disturbances, so as to give us greater latitude in the hypotheses, and (*c*) investigations into the question how far departures from such hypotheses affect the estimation. In other statistical fields such topics have been investigated with considerable profit. There seems no reason to doubt that similar investigations would be rewarding in the theory of structure.

30. To round off this review of the theory of regression and structural analysis I should, perhaps, have mentioned in more detail the use of instrumental variates and problems of multicollinearity connected with Frisch's confluence analysis. My colleague Mr Durbin has, however, been working in this field and I therefore leave these topics in his hands for later treatment.

A first draft of this paper was read by Mr J. Durbin, Mr D. V. Lindley and Dr P. A. P. Moran, to whom I am very grateful for a number of constructive comments and criticisms.

[*Note added in proof.*] In regard to Berkson's proposal (§§ 4–6) there is another possible way of looking at the matter which well illustrates the importance of an unambiguous symbolism and nomenclature in this subject. If we denote observed quantities by a prime, the procedure described in § 5, wherein a value of the variable is aimed at, should be written

$$x_{X'} = X' + \xi_{X'}, \tag{35}$$

where ξ is a variate and hence the latent element $x_{X'}$ is also a variate. The model now asserts that

$$y' = \alpha x_{X'} + \eta, \tag{36}$$

that is to say, the basic relation is not a functional but a structural equation connecting the variates $y', x_{X'}$ and η. If we substitute from (35) we have

$$y' = \alpha X' + (\alpha \xi_{X'} + \eta). \tag{37}$$

The term in brackets on the right is a variate and the estimation of α is then a problem in regression analysis, leading to the least-squares estimation $\Sigma y' X' / \Sigma X'^2$ for α. This is unbiased for fixed X' and hence follows Berkson's result. But it does not prove that there is only one regression or affect the estimation of constants in a functional relationship.

REFERENCES

Anderson, T. W. & Rubin, H. (1949). *Ann. Math. Statist.* **20**, 46.
Berkson, J. (1950). *J. Amer. Statist Ass.* **45**, 160.
Geary, R. C. (1942). *Proc. Roy. Irish Acad.* **47** A, 63.
Geary, R. C. (1943). *Proc. Roy. Irish Acad.* **49** A, 177.
Geary, R. C. (1948). *J.R. Statist. Soc.* B, **10**, 140.
Haavelmo, T. (1944). *Econometrica*, **12**, supplement.
Klein, L. (1950). *Economic Fluctuations in the United States*, 1921–1941. New York: John Wiley and Sons; London: Chapman and Hall.
Koopmans, T. (ed.) (1950). *Statistical Inference in Dynamic Economic Models*. New York: John Wiley and Sons; London: Chapman and Hall.
Koopmans, T. & Reiersöl, O. (1950). *Ann. Math. Statist.* **21**, 165.
Lindley, D. V. (1950). *Proc. Camb. Phil. Soc.* **47**, 337.
Theil, H. (1950). *Indagationes Mathematicae*, **12**, fasc. 2.
Wald, A. (1940). *Ann. Math. Statist.* **11**, 284.

ON THE CONCURRENCE OF A SET OF REGRESSION LINES

By K. D. TOCHER

Imperial College of Science and Technology

Introduction and summary

Many experimental methods determine unknown datum points by plotting two linearly related observables against each other and observing the intercept on one of the axes. It is frequently suspected that the datum changes between experiments, and a test of this is required.

Three cases arise: first, only one observable is subject to error, and the intercept required is on the axis of that observable; secondly, the required intercept is on the axis of the observable free from error; and, lastly, both observables are subject to error.

The first case is solved by a straightforward application of analysis of variance, while the last, although strictly the correct model for all real situations, is beset by many difficulties and, in the absence of any completely satisfactory procedure for solving the ordinary regression problem, is not considered here. Only the second case, which is a close approximation to the situation most frequently arising in this experimental technique, is considered here.

The usual restriction that all errors have equal variance is imposed, but could be relaxed in the case where their ratios are known. A further assumption of normality of error is made, but deviations from this are no more serious than in the usual analysis of variance.

In general, the critical significance level of the test developed is a root of a complicated equation, best obtained by an iterative process described here in detail, but in special cases an exact analytic solution is possible.

If a common intercept does exist (in the sense that the test does not reject that hypothesis) a confidence interval for it may be obtained. The whole process is illustrated by a numerical example, and the extension to composite tests of equality within groups of sets of lines is given.

The problem

Consider a group of m sets of data, the ith set consisting of n_i pairs (x_{ir}, y_{ir}) $(r = 1, 2, \ldots, n_i;$ $i = 1, 2, \ldots, m)$. We assume that there is a linear regression of y on x for each set, viz.

$$E(y_{ir}) = \alpha_i + \beta_i x_{ir}. \tag{1}$$

The x_{ir} are assumed known, and the y_{ir} normally and independently distributed about their mean values with a constant variance of σ^2. The line for the ith set meets the x-axis at a value $x_i = -\alpha_i/\beta_i$. We require to test the hypothesis, H, that $x_i = \text{const.}$ $(i = 1, 2, \ldots, m)$.

The solution

The hypothesis H is a composite one, since the common value for the intercepts is unspecified. Consider one of the constituent simple hypotheses $H(x_0)$ that $x_i = x_0$ $(i = 1, 2, \ldots, m)$, i.e. that

$$\alpha_i + \beta_i x_0 = 0. \tag{2}$$

This is a linear constraint on the unknown parameters α_i, β_i and the usual analysis of variance test can be applied. Put

$$\left.\begin{aligned} \sum_r x_{ir} = n_i \bar{x}_i, \qquad & \sum_r y_{ir} = n_i \bar{y}_i, \\ \sum_r (x_{ir} - \bar{x}_i)^2 = s_i, \quad & \sum_r (y_{ir} - y_i)^2 = S_i, \\ \sum_r (x_{ir} - \bar{x}_i)(y_{ir} - \bar{y}_i) = p_i. & \end{aligned}\right\} \tag{3}$$

The unrestricted sum of squares is

$$S = \sum_i \sum_r (y_{ir} - \alpha_i - \beta_i x_{ir})^2, \tag{4}$$

with minimum value

$$S_0 = \sum_i (S_i - p_i^2/s_i), \tag{5}$$

which has a $\chi^2\sigma^2$ distribution with $\sum_i n_i - 2m = N - 2m$ degrees of freedom. S_0 is obtained from S by substituting the least-square estimates of α_i and β_i which are

$$a_i = \bar{y}_i - b_i \bar{x}_i, \quad b_i = p_i/s_i. \tag{6}$$

The restricted sum of squares obtained by substituting (2) in (4) is

$$S' = \sum_i \{y_{ir} - \beta_i (x_{ir} - x_0)\}^2. \tag{7}$$

The least-squares estimates of β_i in this case are

$$\hat{b}_i = \frac{p_i + n_i(\bar{x}_i - x_0)\,\bar{y}_i}{s_i + n_i(\bar{x}_i - x_0)^2},$$

which reduce (7) to

$$S_0' = \sum_i \left[S_i + n_i \bar{y}_i^2 - \frac{\{p_i + n_i(\bar{x}_i - x_0)\,\bar{y}_i\}^2}{s_i + n_i(\bar{x}_i - x_0)^2} \right].$$

On the hypothesis $H(x_0)$, $S_0' - S_0$ is the sum of squares due to restraints and has a $\chi^2\sigma^2$ distribution with m degrees of freedom, independent of S_0. After some algebraic reduction, we find that

$$S_0' - S_0 = f(x_0) = \sum_i \frac{n_i\{s_i \bar{y}_i - p_i(\bar{x}_i - x_0)\}^2}{s_i\{s_i + n_i(\bar{x}_i - x_0)^2\}} \tag{8A}$$

$$= \sum_i \frac{n_i p_i b_i}{s_i + n_i(\bar{x}_i - x_0)^2}(x_{0i} - x_0)^2, \tag{8B}$$

where x_{0i} is the estimated intercept of the ith line, $-a_i/b_i$.

Thus under $H(x_0)$, $\dfrac{N-2m}{mS_0} f(x_0)$ has the F distribution on m and $N-2m$ degrees of freedom so that this hypothesis is contradicted at the α % significance level if

$$f(x_0) > \frac{mS_0}{N-2m} F_\alpha. \tag{9}$$

The hypothesis H consists of the conjunction of $H(x_0)$ for all x_0. Thus H is contradicted at α % level if (9) holds for all x_0. Conversely, the critical significance level for H is given by the F ratio

$$F_\alpha = \frac{N-2m}{mS_0} \operatorname*{Min}_{x_0} f(x_0). \tag{10}$$

The value of x_0 minimizing $f(x_0)$ is the maximum-likelihood estimate of the common intercept, since $f(x_0)$ only differs from S_0' by the constant S_0. The exact interpretation of the significance level is discussed in the next section.

If, for a given level α, (9) does not hold for all x_0, the exceptional x's form a set of possible values of the intercept, which contain the true common intercept, when it exists, in all but $\alpha\,\%$ of any sequence of independent trials. Thus they form a confidence set for x_0, and since, as we shall see later, $f(x_0)$ is parabolic in shape, this set is the interval between the two roots of

$$f(x) = \frac{mS_0}{N-2m} F_\alpha,$$

and so constitutes a confidence interval for the common intercept.

COMMENTS ON THE SOLUTION

The classical method of determining a test function for a composite hypothesis considers only functions whose probability distributions are independent of the 'nuisance' parameters; in the Neyman-Pearson terminology these functions define similar regions, while the older phrasing is that the appropriate statistic for the corresponding simple hypothesis has been 'Studentized'. From this set of functions one is chosen that has certain optimum properties.

Neyman & Pearson (1933) have shown how to determine such regions under certain conditions, but if those do not hold, there is no general method of procedure known, and for practical purposes we may assume that similar regions do not exist. It is easily shown that the conditions do not hold in this problem of testing equality of intercepts.

Another example of a composite hypothesis with no similar regions is the testing of equality of performance in a comparative trial. In his discussion of such tests Barnard (1947) suggested that the significance level of a test function should be taken as the maximum probability of exceeding that value of the test function, considered as a function of the nuisance parameters. We are then sure that no matter what the true values of these parameters may be, the probability of making the mistake of rejecting the hypothesis when it is true is certainly less than the significance level.

It is in this sense that the criterion suggested for the test of equality of intercepts gives a significance level. The method could be applied to any test function of the constituent simple hypothesis, and the choice of function is governed by the required optimum properties of differentiation from alternative hypotheses. Our test has been based on the maximum-likelihood ratio test which has shown its value as a general method in many instances. Kolodziejczyk (1935) has shown that for the simple hypotheses there are no uniformly most powerful tests, and this indicates that the same will be true for the composite hypothesis.

SPECIAL CASE

In the special case of all sets of data having common values of the x an exact analytic solution is possible.

Put $n_i = n$, $\bar{x}_i = \bar{x}$, $s_i = s$ and $\bar{x} - x_0 = \delta$, then the form (8A) for $f(x_0)$ reduces (9) to the inequality

$$s^2 \sum_i \bar{y}_i^2 - 2s\Sigma\bar{y}_i p_i \delta + \sum_i p_i^2 \delta^2 \geqslant \frac{mS_0 F_\alpha}{n(N-2m)} s(s + n\delta^2). \tag{11}$$

If this is true for all δ we must have

$$\frac{mS_0F_\alpha}{n(N-2m)} = K_\alpha \leqslant \frac{\Sigma p_i^2}{ns}, \tag{12}$$

and $$\sum_i p_i\bar{y}_i^2 - (\sum_i \bar{y}_i^2 - K_\alpha)(\sum_i p_i^2 - nsK_\alpha) \leqslant 0,$$

or on rearranging, that

$$nsK_\alpha^2 - \sum_i (p_i^2 + ns\bar{y}_i^2) K_\alpha + \sum_i \bar{y}_i^2 \sum_i p_i^2 - (\sum_i p_i\bar{y}_i)^2 \geqslant 0. \tag{13}$$

Replacing the inequality in (13) by an equality, the required significance level is obtained from its smaller root, i.e. from

$$F_\alpha = \frac{N-2m}{2mnsS_0}[\sum_i (p_i^2 + ns\bar{y}_i^2) - \{(\sum_i \overline{p_i^2 + ns\bar{y}_i^2})^2 + 4ns(\sum_i p_i\bar{y}_i)^2\}^{\frac{1}{2}}]. \tag{14}$$

In numerical examples, the terms of (14) may be large and nearly equal. In this case it is best to calculate the coefficients of (13) and solve it by a series. If the quadratic is

$$x^2 - ax + b = 0 \quad (0 < b \leqslant a),$$

the lower root is

$$\frac{b}{a}\left\{1 - \frac{b}{a^2} + 2\left(\frac{b}{a^2}\right)^2 \ldots\right\}.$$

The quoted terms are sufficient to give the root, if this method is necessary.

The confidence interval is obtained by rearranging (11) as the quadratic in δ,

$$(\sum_i p_i^2 - nK_\alpha s)\delta^2 - 2s(\sum_i p_i\bar{y})\delta + s^2(\sum_i \bar{y}_i^2 - K_\alpha) = 0, \tag{15}$$

giving two roots δ_1, δ_2 and hence a confidence interval $[\bar{x} - \delta_1, \bar{x} - \delta_2]$ for x_0.

It should be noted that (12) should hold for all this analysis. Writing (12) as

$$\frac{1}{m}\sum_i \frac{p_i^2}{s_i} \Big/ \frac{1}{N-2m}\left(\sum_i S_i - \sum_i \frac{p_i^2}{s_i}\right) \tag{16}$$

we observe that this requires the usual analysis of variance test to detect a difference in slope in the lines at the level α.

If the lines can be assumed to have a common slope, the problem reduces to determining if the lines are identical, a problem soluble by standard techniques.

An iterative solution of the general case

In general, the minimum value of $f(x)$ can be obtained by a lengthy process of interpolation in tables of $f(x)$ and $f'(x)$, but in the majority of cases this may be replaced by an iterative procedure starting from the following simple approximation.

If the variation of estimated intercept from line to line is small, the unknown x_0 in the denominator of each term of (8B) can be replaced by x_{0i} reducing $f(x)$ to a weighted sum of squares

$$f(x) \simeq \sum_i \omega_i(x_{0i} - x)^2, \tag{17}$$

where $$\omega_i = \frac{n_i p_i b_i}{s_i + n_i(\bar{x}_i - x_{0i})^2}.$$

This approximation produces a minimum value

$$\min_x f(x) \doteq \Sigma\omega_i x_{0i}^2 - \frac{(\Sigma\omega_i x_{0i})^2}{\Sigma\omega_i} \tag{18}$$

at the x value $\dfrac{\Sigma\omega_i x_{0i}}{\Sigma\omega_i}$.

It is of interest to note that the weights used in this weighted sum of squares are those obtained by an application of the well-known device to obtain the variances of the individual intercepts:*

$$x_{0i} = \bar{x}_i - \frac{\bar{y}_i}{b_i}, \quad V(x_{0i}) \doteq \frac{1}{b_i^2} V(\bar{y}_i) + \bar{y}_i^2 V\left(\frac{1}{b_i}\right)$$

$$= \frac{\sigma^2}{b_i^2}\left[\frac{1}{n_i} + \frac{\bar{y}_i^2}{b_i^2}\frac{1}{s_i}\right]$$

$$= \frac{s_i + n_i(\bar{x}_i - x_{0i})^2}{n_i p_i b_i}\sigma^2 = \frac{\sigma^2}{\omega_i}.$$

The iteration is required to obtain an exact solution to the equation

$$f'(x) = 2\sum_i \frac{n_i p_i b_i\{s_i + n_i(\bar{x}_i - x)(\bar{x}_i - x_{0i})\}}{\{s_i + n_i(\bar{x}_i - x)^2\}^2}(x_{0i} - x) = 0, \tag{19}$$

and a suitable one is obtained if we put

$$x^{(n)} = \frac{\Sigma\omega_i^{(n)} x_{0i}}{\Sigma\omega_i^{(n)}},$$

where

$$\omega_i^{(n+1)} = \frac{n_i p_i b_i\{s_i + n_i(\bar{x}_i - x^{(n)})(\bar{x}_i - x_{0i})\}}{\{s_i + n_i(\bar{x}_i - x^{(n)})^2\}^2}, \tag{20}$$

starting from the initial values

$$\omega_i^{(0)} = \omega_i.$$

If (20) converges, it clearly converges to a root of (19), and this root, being a weighted mean of x_{0i}, lies in the range of x_{0i}.

To investigate the convergence of the iteration we consider the solution of the equation $x = \psi(x)$ by the iteration $x_{n+1} = \psi(x_n)$. We have for a root x

$$x_{n+1} - x = \psi(x_n) - x = \psi(x_n) - \psi(x) = (x_n - x)\psi'(\xi),$$

where ξ lies in the range $[x_n, x]$. Thus if $\psi'(\xi) < 1$ for all ξ in the neighbourhood of the root x, and x_0 is chosen in that neighbourhood, the iteration will converge at least as fast as a geometrical progression. This is only a sufficient condition. Necessary conditions are more involved. If $\psi'(x)$ is continuous, then it suffices if $\psi'(x) < 1$ at the root and x_0 is near enough.

In the application

$$\psi(x) = \frac{\Sigma\omega_i(x)\, x_{0i}}{\Sigma\omega_i(x)},$$

where

$$\omega_i(x) = n_i p_i b_i \frac{s_i + n_i(\bar{x}_i - x)(\bar{x}_i - x_{0i})}{\{s_i + n_i(\bar{x}_i - x)^2\}^2}.$$

We easily obtain $\psi'(x)$ at the root $x = x_0$,

$$\psi'(x_0) = \frac{\sum_i \omega_i'(x_0)(x_{0i} - x_0)}{\Sigma\omega_i(x_0)},$$

* I am indebted to E. C. Fieller for pointing out this interesting fact.

and
$$\omega_i'(x_0) = \omega_i(x_0)\left\{\frac{x_{0i}-\bar{x}_i}{s_i+n_i(\bar{x}_i-x_0)(\bar{x}_i-x_{0i})} - \frac{4(x_0-\bar{x}_i)}{s_i+n_i(\bar{x}_i-x_0)^2}\right\} \tag{21}$$
$$\doteq -\frac{3(x_{0i}-\bar{x}_i)}{s_i+n_i(\bar{x}_i-x_{0i})^2}\omega_i(x_0) = -\mu_i\omega_i(x_0) \text{ say.}$$

Note that μ_i is independent of x_0 and can be calculated direct from the data.

An approximate sufficient condition is

$$\left|\sum_i \omega_i(x_0)\mu_i(x_{0i}-x_0)\right| \leqslant \sum_i \omega_i(x_0),$$

since
$$\omega_i(x_0) > 0 \quad (i = 1, 2, \ldots, m),$$

which certainly holds if

$$|\mu_i| < \frac{1}{|x_{0i}-x_0|} \quad (i = 1, 2, \ldots, m). \tag{22}$$

Inserting the first approximation for x_0 in (22) gives approximate sufficient conditions for convergence. They can be compounded into a weaker condition by noticing that as x_0 lies in range R of the x_{0i}, (22) is satisfied if

$$R = \operatorname*{Max}_i x_{0i} - \operatorname*{Min}_i x_{0i} < \frac{1}{\operatorname*{Max}_i |\mu_i|}. \tag{23}$$

The minimum value of $f(x)$ is obtained as

$$\Sigma\lambda_i(x)(x_{0i}-x)^2, \tag{24}$$

where
$$\lambda_i(x) = \frac{n_i p_i b_i}{s_i+n_i(x-\bar{x}_i)^2}, \tag{25}$$

and x is the exact root.

Approximate confidence limits can be derived from the approximate form (17) of $f(x)$ as the roots of

$$\sum_i \omega_i x^2 - 2\left(\sum_i \omega_i x_{0i}\right)x + \left(\sum_i \omega_i x_{0i}^2 - \frac{mS_0}{N-2m}F_\alpha\right) = 0. \tag{26}$$

The exact equation to be solved is

$$\left\{\sum_i \lambda_i(x)\right\}x^2 - 2\left\{\sum_i \lambda_i(x)x_{0i}\right\}x + \left\{\sum_i \lambda_i(x)x_{0i}^2 - \frac{mS_0F_\alpha}{N-2m}\right\} = 0,$$

or, using this as a definition of the W's, as

$$W_0(x)x^2 - 2W_1(x)x + W_2(x) = 0. \tag{27}$$

This can be solved using the iteration

$$x_{n+1} = W_0(x_n)x_n^2 - \frac{W_2(x_n)}{2\{W_0(x_n)x_n - W_1(x_n)\}} \tag{28}$$

for each root in turn, using those of (26) as initial values.

The detailed investigation of convergence is rather complicated, but a simple account can be given neglecting the change of the W's with x. In this case the iteration is simply

$$x_{n+1} = \frac{W_0x_n^2 - W_2}{2(W_0x_n - W_1)}.$$

This has a ψ function

$$\psi(x) = \frac{W_0x^2 - W_2}{2(W_0x - W_1)}, \quad \psi'(x) = \frac{W_0(W_0x^2 - 2W_1x + W_2)}{2(W_0x - W_1)^2}.$$

$\psi'(x)$ vanishes at the root and so the iteration converges very rapidly.

A NUMERICAL EXAMPLE

In certain colour-temperature experiments on electric filament lamps, the power of a lamp is obtained from the response of a photoelectric cell to the illumination from the lamp for varying lamp to cell distances; the root reciprocal response is a linear function of this distance by the inverse square law of illumination. The 'zero' point of the cell is not known and is estimated from the intercept of the regression line. It is suspected that the zero point changes with the type of lamp or the frequency of light filtered to the cell. The data below are from one experiment of this series using the same lamp at different distances with two different filters:

Lamp distance x (cm.)	Root reciprocal response	
	y_1 Filter 1	y_2 Filter 2
2278	0·802	0·739
2406	0·882	0·812
2524	0·956	0·879
2638	1·028	0·944
2760	1·101	1·012

Both lamp 'zero' about 1000 cm., and the large extrapolation involved causes the arithmetic to be rather heavy. The notation, exemplified by $7{\cdot}139^{-4} = 7{\cdot}139 \times 10^{-4} = 0{\cdot}0007139$, is adopted throughout. We easily obtain

$$\Sigma x = 1{\cdot}2606^{4}, \quad \sum_r \bar{y}_{1r} = 4{\cdot}769^{0}, \quad \sum_r y_{2r} = 4{\cdot}386^{0},$$

$$\bar{x} = 2{\cdot}5212^{3}, \quad \bar{y}_1 = 0{\cdot}9538^{0}, \quad \bar{y}_2 = 0{\cdot}8772^{0},$$

$$s = 1{\cdot}43093^{5}, \quad S_1 = 5{\cdot}53768^{-2}, \quad S_2 = 4{\cdot}59868^{-2},$$

$$p_1 = 8{\cdot}90132^{1}, \quad p_2 = 8{\cdot}11188^{1},$$

$$b_1 = 6{\cdot}22066^{-4}, \quad b_2 = 5{\cdot}66896^{-4},$$

$$x_{01} = 9{\cdot}87923^{2}, \quad x_{02} = 9{\cdot}73828^{2},$$

$$S_0 = 5{\cdot}53386^{-6}.$$

Hence

$$ns = 7{\cdot}15464^{5}, \quad \Sigma p^2 \Sigma \bar{y}^2 - (\Sigma p \bar{y})^2 = 5{\cdot}05905^{-1},$$

$$\Sigma p \bar{y} = 1{\cdot}56058^{2}, \quad \Sigma p^2 + ns \Sigma \bar{y}^2 = 1{\cdot}215921^{6}.$$

Equation (13) takes the form

$$7{\cdot}15464^{5} K^2 - 1{\cdot}215921^{6} K + 5{\cdot}05905^{-1} = 0,$$

or

$$K^2 - 1{\cdot}699485\, K + 7{\cdot}071003^{-7} = 0,$$

i.e.

$$K = 4{\cdot}1607^{-7}\{1 + O(10^{-7}) \ldots\},$$

$$F = \frac{5 \times 6 \times 4{\cdot}1607^{-7}}{2 \times 5{\cdot}53386^{-6}} = 1{\cdot}128.$$

Thus the difference of intercepts is not significant.

Confidence limits on x are given by (15).

From tables of the F distribution:

$$F_{0\cdot05} = 5\cdot14, \quad K_{0\cdot05} = \frac{2\times5\cdot53386^{-6}}{5\times6} = 1\cdot896^{-6},$$

$$\delta = \frac{S}{\Sigma p_i^2 - nsK_{0\cdot05}}[(\Sigma p\bar{y}) \pm \{(\Sigma py)^2 - (\Sigma p)(\Sigma y) + (\Sigma p^2 + ns\Sigma y^2)K_{0\cdot05} - nsK_{0\cdot05}^2\}^{\frac{1}{2}}]$$

$$= \frac{1\cdot43093^5}{1\cdot45023^4}[1\cdot56058^2 \pm \{-5\cdot05905^{-1} + 2\cdot30571^0 - 2\cdot57\ldots^{-6}\}^{\frac{1}{2}}]$$

$$= 9\cdot86694[1\cdot56058^2 \pm \sqrt{(1\cdot79981)}]$$

$$= 1\cdot53982^3 \pm 1\cdot3237^1,$$

$$x_0 = 9\cdot6815^2 \quad \text{or} \quad 9\cdot9462^2.$$

Extension of analysis to a classified set of regression lines

If a significantly large variation of intercept is found by this test in data which are divided into groups, it is natural to ask whether the variation is only between the groups, with a common intercept in each group.

Let there be p groups of sets, the ith group consisting of q_i sets, the jth set of which has n_{ij} pairs (x_{ijr}, y_{ijr}) with a linear regression

$$E(y_{ijr}) = \alpha_{ij} + \beta_{ij}x_{ijr} \quad (r = 1, 2, \ldots, n_{ij};\ j = 1, 2, \ldots, q_i;\ i = 1, 2, \ldots, p), \tag{29}$$

and the same assumptions concerning the x's and y's as in the simpler problem.

The composite hypothesis H' is that there are p numbers x_{01}, x_{02}, x_{0p} such that

$$\alpha_{ij} + \beta_{ij}x_{0i} = 0 \quad (j = 1, 2, \ldots, q_i;\ i = 1, 2, \ldots, p). \tag{30}$$

A constituent simple hypothesis $H'(x_{01}, x_{02}, \ldots, x_{0p}) = H'(x_{0i})$ is that (30) is true for a given set $x_{01}, \ldots, x_{0p}$.

Using an obvious extension of the notation of (3) the minimized unrestricted sum of squares is

$$S_0 = \sum_i \sum_j (S_{ij} - p_{ij}^2/s_{ij}), \tag{31}$$

while the sum of squares due to restraints can be shown to be

$$\sum_i \sum_j \frac{n_{ij}p_{ij}b_{ij}(x_{0i} - x_{0ij})^2}{s_{ij} + n_{ij}(x_{0i} - \bar{x}_{ij})^2} = \sum_i f_i(x_{0i}), \tag{32}$$

where b_{ij}, x_{0ij} are the estimated slope and intercept of the jth line of the ith group.

These expressions are distributed as $\chi^2\sigma^2$ with

$$\sum_i \sum_j (n_{ij} - 2) = N - 2\sum_i q_i = N - 2Q$$

and Q degrees of freedom respectively, and thus the significance level is given by

$$F_\alpha = \frac{N-2Q}{QS_0}\operatorname*{Min}_{x_{0i}} \sum_i f_i(x_{0i}) = \frac{N-2Q}{QS_0}\sum_i \operatorname*{Min}_x f_i(x),$$

which has the F distribution with Q and $N - 2Q$ degrees of freedom.

The numerical processes described above can be applied to determine F_α. The minimizing values of x_{0i} are maximum likelihood estimates of the common intercepts, but if confidence intervals are required, these can best be obtained from each group separately.

In a normal analysis of variance the difference between the two sums of squares calculated, assuming, first, one common intercept and, secondly, one for each group, has a $\chi^2\sigma^2$ distribution and can be used as a direct test of the significance of between-group variations. In this case the two sums of squares of a constituent simple hypothesis have the same degrees of freedom, and the set of simple hypotheses contained in H is a subset of H'.

The further extension of the problem to multiple classification follows the lines already laid down and needs no further elaboration.

REFERENCES

BARNARD, G. A. (1947). Significance tests for 2×2 tables. *Biometrika*, **34**, 123.

KOLODZIEJCZYK, ST (1935). On an important class of statistical hypotheses. *Biometrika*, **27**, 161.

NEYMAN, J. & PEARSON, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans.* A, **231**, 289.

A SAMPLING TEST OF THE χ^2 THEORY FOR PROBABILITY CHAINS

By M. S. BARTLETT

University of Manchester

In my paper on the χ^2 theory for probability chains (Bartlett, 1951) it was stated that numerical examination of the proposed methods would be reported later. The purpose of the present note is to demonstrate numerically the relevance of the theory in the case of a simple stationary Markoff chain with two possible states. Two separate cases were considered with transition probabilities from one state to the other given by the columns of the matrices in Table 1. These transition probabilities are easily seen to give the following

Table 1

	Case I Initial state			Case II Initial state	
	0	1		0	1
Final state 0	1/3	2/3	Final state 0	2/3	1/2
Final state 1	2/3	1/3	Final state 1	1/3	1/2
	1	1		1	1

expected relative frequencies of consecutive pairs of states in a long sequence (Table 2). It was shown in my paper (p. 89) that χ^2 theory may be applied, at least for sufficiently long sequences, to the two-way frequencies of Tables 1 and 2, but not (in contrast with the complete independence case) to the marginal frequencies alone. In case I the expected value of χ^2 for the marginal frequencies was given as 0·5, in case II as 1·4 (for 1 nominal degree of freedom).

Table 2

	Case I Initial state			Case II Initial state		
	0	1		0	1	
Final state 0	1/6	1/3	1/2	2/5	1/5	3/5
Final state 1	1/3	1/6	1/2	1/5	1/5	2/5
	1/2	1/2	1	3/5	2/5	1

A sequence of 1000 values was obtained from the random numbers in the tables of Fisher and Yates's, by arranging for the first value to occur according to the marginal probability,

and the succeeding values to occur with the transition probabilities appropriate to each previous state. This was done for both cases I and II (and subsequently repeated for reasons given below). Each sequence was subdivided into ten subsequences of 100, thus giving rise to ten 2×2 tables. For example, the first 100 values obtained in case I were:

0 0 0 1 0, 1 1 0 0 1, 0 1 0 1 0, 1 0 0 1 0, 1 0 0 1 0,
0 1 0 1 0, 1 0 1 1 1, 0 1 1 0 0, 0 1 1 0 1, 0 1 0 0 0,
0 1 1 0 0, 0 1 1 0 1, 0 1 1 1 1, 0 1 1 1 0, 1 1 0 0 1,
0 1 0 1 0, 1 1 0 1 1, 0 0 1 1 1, 1 0 1 0 1, 1 1 0 1 0.

Table 3

	0_1	1_1	
0_2	15	32	47
1_2	32	20	52
	47	52	99

Table 3 shows the corresponding 2×2 table. (Since the total for 0_1 does not include the final value, which is 0, and the total for 0_2 the first value, which is 0, both totals in this sample agree at 47, one less than the total number of 0's.)

The expected frequencies for a total sample of 99 are obtained from Table 2 by multiplying the probabilities given there by this total frequency. A χ^2 could then be worked out, the standard quadratic expression (no corrections for continuity) in the deviations of observed from expected frequencies being used. This value for the two-way frequencies is denoted by χ_1^2. The similar value for the marginal frequencies (for definiteness, when the row and column values differed by unity, the column totals were used) is denoted by χ_2^2, and an appropriate χ^2 expression for the two-way frequencies is then

$$\chi^2 = \chi_1^2 - \chi_2^2 \text{ with 2 degrees of freedom.}$$

For Table 3 this gives $\chi^2 = 0{\cdot}938 - 0{\cdot}252 = 0{\cdot}686.$

The complete set of values so obtained is given in Table 4. The average values are shown in brackets at the foot of Table 4, and it will be seen that, in spite of the usual fluctuations in individual values, these average values seem to agree reasonably well with the expected values (for χ_2^2 in case I, 0·5; in case II, 1·4; for χ^2 in either case, 2·0). The total of all χ^2's in case I is 37·66 (40 D.F.) and in case II 51·80, both values being quite admissible. On the other hand, the total of χ_2^2's in case I is 9·32 (nominally 20 D.F.), and in case II 40·82. The former is significantly low at the 0·05 level, and the latter significantly high at the 0·01 level.

Although the asymptotic expected value of χ_2^2 is known, its distribution is not. However, we might as a further test demonstrate that while the values of χ^2 are not significantly different for cases I and II, those of χ_2^2 are, clustering round 0·5 and 1·4 respectively. It seemed advisable from the nature of the distributions to carry out these tests on the square-root scale, in order to reduce the effect of skewness. In testing a difference, not between two observed means, but between a mean and a theoretical value, it also seemed advisable to use the correct mean on the square-root scale, which is

$$\sqrt{2}\,\Gamma\tfrac{1}{2}(n+1)/\Gamma\tfrac{1}{2}n$$

for a χ^2 distribution with n degrees of freedom. For χ_2^2 it was assumed that the correction factor 0·798 appropriate to the nominal number of degrees of freedom (namely, one) could be used; the test is not of course very dependent on this particular assumption. This gives expected values on the square-root scale of 1·253 for χ^2, 0·564 for χ_2^2 in case I and 0·945 in case II. The relevant t statistics came out as follows:

Case I, χ^2 against theoretical mean	$t(19\text{ D.F.}) = -0{\cdot}693$
Case II, χ^2 against theoretical mean	$t = 1{\cdot}938$
Case I versus case II	$t(38\text{ D.F.}) = -1{\cdot}745$
Case I, χ_2^2 against theoretical mean	$t(19\text{ D.F.}) = -0{\cdot}011$
Case II, χ_2^2 against theoretical mean	$t = 1{\cdot}222$
Case I versus case II	$t(38\text{ D.F.}) = -3{\cdot}190.$

Table 4

Case I			Case II		
χ_1^2	χ_2^2	χ^2	χ_1^2	χ_2^2	χ^2
0·938	0·252	0·686	4·581	1·833	2·748
1·484	0·252	1·232	10·949	5·470	5·479
3·667	0·252	3·415	3·475	0·750	2·725
5·637	0·818	4·819	0·545	0·285	0·260
0·999	0·090	0·909	5·394	0·108	5·286
8·182	1·708	6·474	0·343	0·082	0·261
1·999	0·494	1·505	2·968	0·890	2·078
0·908	0·090	0·818	9·838	5·663	4·175
11·667	2·920	8·747	1·228	0·545	0·683
1·999	0·494	1·505	7·287	5·663	1·624
2·394	0·252	2·142	0·546	0·108	0·438
0·454	0·090	0·364	5·666	0·815	4·851
1·999	0·818	1·181	1·455	0·545	0·910
0·938	0·252	0·686	5·324	2·305	3·019
0·030	0·010	0·020	1·176	0·007	1·169
2·727	0·252	2·475	8·399	4·729	3·670
0·272	0·090	0·182	10·419	5·470	4·949
0·181	0·090	0·091	10·418	5·470	4·948
0·060	0·010	0·050	0·243	0·007	0·236
0·454	0·090	0·364	2·364	0·082	2·282
(2·350)	(0·466)	(1·883)	(4·631)	(2·041)	(2·590)

The last value of t is highly significant, as anticipated. It should be mentioned that at first only the ten sets of 100 making the first sequence of 1000 in each case were considered, and the corresponding value of t at that stage was 2·405 (18 D.F.). It was felt that this value, though significant, might not be sufficiently convincing, and so a second sequence of 1000 in each case was added. This sequential decision somewhat vitiates the nominal significance level for the final t value, but the evidence for the anticipated difference in χ_2^2 in cases I and II (the difference being in the correct direction) seems clear enough.

The rather high, though insignificant (insignificant also for the first sequences alone),

values of t occurring in the test of χ^2 are mainly due to the total χ^2 in case I exceeding expectation, and it will be remembered that the more exact direct test of the total χ^2 gave less cause for suspicion, giving an equivalent normal fluctuation (of unit standard deviation) of 1·27. It is, of course, possible that in *finite* samples the values of χ^2 are somewhat less restricted when χ_2^2 has a larger average value; to settle this point, a more extensive investigation would be required.

I am indebted to Mrs G. W. Walls for assistance in this investigation.

REFERENCE

BARTLETT, M. S. (1951). The frequency goodness of fit test for probability chains. *Proc. Camb. Phil. Soc.* **47**, 86.

ON MATHEMATICAL ANALYSIS OF STYLE

By WILHELM FUCKS, *Aachen*

1. Introduction

Every significant text of a grammatical exposition consists of a certain material, the *vocabulary*, and some structural properties, the style, of its author. The passive vocabulary is formed by the totality of all words of that language, s, the author writes in, the active vocabulary is formed by a certain set, s', of that totality, the selection of which is determined essentially by the sort of literature the text belongs to and depends only in a lower degree on the peculiarity of the author. Style, however, is characteristic of the author at a certain period of his personal development.

The aim of the following investigation is to formulate mathematically some of the properties of structure constituting style, so that for a given text the application of a simple mathematical criterion allows its attribution to a particular author at a certain period of his mental development.

2. The methodology of style analysis

The analysis of the structural properties of language can be performed in two principal ways:

(1) *synthetically*, by analysing the formation of a text and its structure by means of its elements;

(2) *analytically*, by analysing a given text as to the structural properties of its elements.

The synthetic method has gained considerable importance in communication engineering through the 'information theory'. It represents the so-to-speak 'atomistic' foundation of the more 'phenomenologic' methods of the analytical treatment which constitutes the proper field of style analysis. The relation between the two methods is analogous to the relation between the methods of statistical mechanics and thermodynamics, and it will be seen that this analogy is not only formal, but has a deeper significance.

3. The analytical investigation of language

The elements of the totality of significant grammatical manifestations can be ordered in the following sequence, each element of which contains the preceding one as a sub-element:

letter
syllable
word
sentence
paragraph
chapter
book
work
total work of an author
kind of literature within a language
total literature of a language
kind of literature
world literature

The text of a grammatical manifestation can be conceived as a linear set, the elements of which may be numbered by a position number n. Now we consider functions $f(n)$ defined on this set and seek for relations between the functional values representing characteristic structural properties of the text. Such functions are, for example, the number of letters or syllables of the nth word, the number of words in the nth sentence, the metrum, and so on.

As numerical characteristics of the relations between the functional values, we have to consider the average number of letters or syllables per word, the relative frequencies of the one-syllabled, two-syllabled, etc., words, the average distances between an i-syllabled and a k-syllabled word and so on. Which of these numerical characteristics will be peculiar to style can only be deduced from a more profound inquiry based on a new conception of text structure.

We distinguish, for the moment only qualitatively, prose and poetic language. Between these extremes there are continuous transitions. We will try to comprehend quantitatively the 'binding character' by a numerical characteristic. By an analogy with physics we can regard the states of the text-elements *before* the formation of text as 'gaseous', in prose-text as 'linearly fluid' and in poetic text as 'linearly crystalline'. The formation process of text then corresponds with condensation (transition gaseous-fluid in case of prose, gaseous-solid in case of poetic text, respectively). This analogy brings out which mathematical aids of style analysis are to be applied in a particular case. The properties of symmetry and periodicity of the 'crystalline' state make group theory the appropriate instrument; and, indeed, it has already been brought into style-analytical investigations (cf. Bense, 1949). The 'gaseous' and 'fluid' states, however, are only accessible to a statistical or phenomenologic-thermodynamic treatment. Statistical methods of style analysis using *words* as elements have been developed by G. U. Yule (1939, 1944). As numerical characteristics specific for style he gives the average number of words per sentence (average sentence lengths), and the frequency distribution of substantives, i.e. the number of substantives occurring once, twice, and so on, respectively.

4. THE SYNTHETIC INVESTIGATION OF LANGUAGE

We regard the formation process of a text as an ordered arrangement of language elements selected in a certain way from a stock, V, of elements. As language elements there must be taken into consideration: letters (V = alphabet), syllables (V = syllable stock), words (V = vocabulary), sentences (V = sentence set). We suppose the stock to contain a sufficient number of samples of each element. While the statistical investigations of language have so far employed words as elements, we will here choose *syllables* as the natural elements of language. In communication engineering the terms 'Sprachatome' or 'Logatome' are used. It is well known that communication engineering utilizes these natural elements of speech-flux to define the 'Silbenverständlichkeit' as a measure for the efficiency of a transmitting system. Now we imagine the words a_k^i occurring in the vocabulary V as balls, marked with the syllable number i, contained in the urn V with certain relative frequencies p_i of the marks i. The formation of a text then represents a certain succession of drawings of the balls a_k^i out of the urn V and composition in form of a sequence

$$a_1^i a_2^j a_3^k \dots a_n^l \dots \begin{cases} i, j, \dots l, \dots & = \text{syllable numbers,} \\ 1, 2, \dots, n, \dots & = \text{position numbers of the words } a. \end{cases}$$

The drawings and consequently the succession of the words or their marks in the text are, however, in general not stochastically independent, but there exist conditioned probabilities for the occurrence of certain configurations of words:

$$w(a_r^i, a_s^j), \quad w(a_r^i, a_s^j, a_t^k),$$

and so on, and certain configurations of marks:

$$w(i,j), \quad w(i,j,k),$$

and so on, respectively. These very probabilities represent characteristic structural properties of the text. Conversely, if these probabilities are given for a stock V of elements with definite relative frequencies, the drawing process described above produces an artificial language with the prescribed structural properties. This synthesis of language means therefore a stochastic process in the form of a 'Markoff-Kette'.

Examples for a sequence of such artificially produced languages, which approach the English language, are given by C. E. Shannon (1948).

The synthetical investigation of language gives a hint as to the direction in which characteristics of style may be found analytically. The consideration of conditioned probabilities for the occurrence of polynomial sequences of elements is not possible because of the difficulty of computation. Thus the synthetical method can generate artificial languages with 'Nahordnung' only. For style analysis, however, the very relations of connexion and succession of text elements over a wider range are important. In order to characterize them, there are (*inter alia*) the following possibilities:

(1) the description of *correlation* between the marks at different positions of the text by means of an appropriate measure of correlation,

(2) the estimation of *average distances* between two equal or different marks in the text.

In our thermodynamic analogy the word-stock V, or the text, corresponds to a gas mixture or a fluid mixture (prose) or a mixed crystal (poetic form), respectively, consisting of components with different marks. The 1-syllabled, 2-syllabled, ..., words correspond perhaps to 1-atomic, 2-atomic, ..., molecules. They can exist in the text in different states of order, they can be distributed at random or only possess 'Nahordnung', or perhaps be ordered over a wider range in a superstructure. By this means the binding character of language will be shown by the extent of a superstructure such as this, which in turn can be ascertained by estimating the correlation between the components at different positions as well as by determining the average distances—that is, the mean paths in the linear fluid or the average lattice constant in the linear crystal.

Further, the order-disorder relations can be comprehended quantitatively by an appropriately defined *entropy-function*.

5. Computation of style-characteristics

(5·1). *The syllable measures*

First of all we determine the *distribution* of syllable numbers, that is to say, the relative frequencies p_i of the i-syllabled words. Thus, if A_i denotes the number of the i-syllabled words and N the total number of words,

$$p_i = \frac{A_i}{N}. \tag{1}$$

By counting, the values shown in Table 1 were obtained.

Table 1. *Percentage distribution of i-syllabled words*

		$100 \times p_i$ (%)					
Author	Work	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$
Shakespeare	*Othello*	78·81	15·11	4·95	1·16	—	—
Galsworthy	*Swan Song I*	75·18	17·55	5·58	1·39	—	—
	Swan Song II	75·54	17·40	5·93	1·13	—	—
	Forsyte Saga	73·44	18·36	6·06	1·48	—	—
Huxley	*Brave New World*	69·82	19·64	6·96	2·43	—	—
	Antic Hay	68·68	19·02	7·93	2·61	—	—
Rilke	*Cornet*	62·34	30·30	5·99	1·34	—	—
Carossa	*Geheimnisse des reifen Lebens*	50·13	31·94	11·99	4·35	1·17	0·35
Hesse	*Steppenwolf*	51·81	31·27	11·22	4·29	1·08	0·26
Mann	*Buddenbrooks*	48·60	33·00	12·20	4·79	1·14	0·27
	Zauberberg	52·91	29·00	11·62	4·38	1·37	0·37
Jaspers	*Der philosophische Glaube*	50·54	25·04	12·40	7·45	3·24	0·79

It can be seen from this table that an increase in the percentage of many-syllabled words is associated with an increase in the prose character of the text. This phenomenon will affect the following numerical characteristics, all of which are computed from the p_i.

The *average syllable number per word* is defined by

$$\bar{\imath} = \sum_i A_i/N = \sum_i i \,.\, p_i. \tag{2}$$

Values of $\bar{\imath}$ are given in Table 2.

Table 2. *Average syllable number per word*

Author	Work	$\bar{\imath}$	Author	Work	$\bar{\imath}$
Shakespeare	*Othello*	1·287	Rilke	*Cornet*	1·464
Galsworthy	*Swan Song I*	1·326	Carossa	*Geheimnisse*	1·732
	Swan Song II	1·338	Hesse	*Steppenwolf*	1·721
	Forsyte Saga	1·342	Mann	*Buddenbrooks*	1·807
Huxley	*Brave New World*	1·397		*Zauberberg*	1·723
	Antic Hay	1·409	Jaspers	*Der philos. Glaube*	1·885

By this means we ensure, in contrast to the relative frequencies, that a particular work of a particular author is associated with a single characteristic number. Furthermore, this average syllable number per word is peculiar to the author.

(5·2). *The average distances*

The average distance l_{ik} between an i-syllabled and a k-syllabled word is defined by the mean value of the differences of position numbers between an i-syllabled and the next k-syllabled word.

For example, the matrix of mean paths for Rilke's *Cornet* reads:

$$l_{ik} = \begin{pmatrix} 1{\cdot}605 & 1{\cdot}081 & 1{\cdot}485 & 1{\cdot}594 \\ 1{\cdot}065 & 3{\cdot}422 & 3{\cdot}113 & 3{\cdot}704 \\ 1{\cdot}476 & 2{\cdot}933 & 16{\cdot}83 & 14{\cdot}38 \\ 1{\cdot}515 & 4{\cdot}111 & 13{\cdot}72 & 77{\cdot}13 \end{pmatrix}.$$

Since we see that the diagonal elements show mutually the most relevant differences we shall use them only, not the whole matrix. Moreover, this brings considerable advantage in computation of the average distances, because of the following simple relation for the relative frequencies p_i.

The syllable number i is a function, $i = f(n)$, of the word position n. The distance between two succeeding i-syllabled words is defined by

$$f(n + l_{ii}^{(1)}) = f(n) = i. \tag{3}$$

Further
$$f(n + l_{ii}^{(1)} + l_{ii}^{(2)} + \ldots) = f(n) = i. \tag{4}$$

If N is the total number of words and A_i the number of i-syllabled words contained in them, if n_a is the number of words before the first i-syllabled, and n_e the number of words after the last i-syllabled word (up to N), then

$$n_a + l_{ii}^{(1)} + l_{ii}^{(2)} + \ldots + l_{ii}^{(A_i - 1)} + n_e = N - 1. \tag{5}$$

The *mean* path is defined by

$$\bar{l}_{ii} = \frac{1}{A_i - 1} \sum_{\alpha=1}^{A_i - 1} l_{ii}^{(\alpha)} = \frac{N - (1 + n_a + n_e)}{A_i - 1}$$

$$\cong \frac{N}{A_i} \quad \text{for } N \to \infty \text{ and } p_i = \text{fixed}. \tag{6}$$

Therefore
$$(\bar{l}_{ii})_\infty = \frac{1}{p_i}. \tag{7}$$

This relation makes it possible to compute directly from the relative frequencies the mean distances between equal-syllabled words, that is to say, the diagonal elements of the l_{ik}-matrices. Table 3 shows the values obtained.

These results have the advantage of exhibiting, in contrast to the relative frequencies, the characteristic differences in the distribution of the poly-syllabled words.

Hence we can deduce another numerical characteristic, the *'trace' or 'spur' of the l_{ik}-matrix*

$$s = \sum_i \bar{l}_{ii}. \tag{8}$$

We can only make comparisons as far as we have calculated the average distances. Thus we carry out the summation in all cases up to $i = 4$ (see Table 4).

This trace, which is introduced to get again a single characteristic from the sequence of numerical characteristics of the mean paths, is also a very specific one, for its values for the different works of one author lie close together.

Table 3. *Average distances between equal-syllabled words*

	Rilke	Carossa	Hesse	Mann		Jaspers
	Cornet	*Geheimnisse*	*Steppenwolf*	*Buddenbrooks*	*Zauberberg*	*Phil. Glaube*
l_{11}	1·604	1·995	1·930	2·058	1·890	1·979
l_{22}	3·300	3·130	3·198	2·897	3·449	3·994
l_{33}	19·28	8·340	8·914	8·189	8·604	8·063
l_{44}	74·10	23·01	23·34	20·90	22·83	13·43
l_{55}	—	85·40	92·42	87·95	73·21	30·86
	Shakespeare	Galsworthy			Huxley	
	Othello	*Swan S. I*	*Swan S. II*	*Forsyte S.*	*Brave N. W.*	*Antic Hay*
l_{11}	1·269	1·330	1·324	1·362	1·432	1·456
l_{22}	6·617	5·698	5·747	5·447	5·094	5·259
l_{33}	20·21	17·94	16·86	16·51	14·38	12·61
l_{44}	79·55	71·84	70·13	67·75	41·14	38·39

Table 4. *Trace of the mean-path matrices*

Author	Work	Spur s	Author	Work	Spur s
Shakespeare	*Othello*	107·65	Rilke	*Cornet*	99·09
Galsworthy	*Swan Song I*	96·80	Carossa	*Geheimnisse*	36·48
	Swan Song II	94·06	Hesse	*Steppenwolf*	37·38
	Forsyte Saga	91·07	Mann	*Buddenbrooks*	34·04
Huxley	*Brave New World*	62·04		*Zauberberg*	36·78
	Antic Hay	57·72	Jaspers	*Der Philos. Glaube*	27·47

(5·3). *Entropy*

An especially characteristic measure for the order-disorder relations in a text should be the statistical entropy. It is defined by

$$S = -k \sum_i p_i \log p_i, \tag{9}$$

the p_i being relative frequencies and k an arbitrary constant. In our thermodynamic conception of text as a mixture of components i, with the frequencies p_i, S just represents the 'mixture entropy'. This entropy expression holds for independent text elements, i.e. systems 'without interrelation'. If we wish to consider the interrelations, we have to form

$$S = -k \sum_i \sum_k p_{ik} \log p_{ik}, \text{ etc.}, \tag{10}$$

where p_{ik} denotes the relative frequencies of pairs: i, k.

With the values of Table 1, the entropy values given in Table 5 are computed (S was calculated with the aid of Briggs's logarithms).

Table 5. *Entropy values*

Author	Work	Entropy S	Author	Work	Entropy S
Shakespeare	*Othello*	0·2940	Rilke	*Cornet*	0·3836
Galsworthy	*Swan Song I*	0·3215	Carossa	*Geheimnisse*	0·4783
	Swan Song II	0·3233	Hesse	*Steppenwolf*	0·4710
	Forsyte Saga	0·3344	Mann	*Buddenbrooks*	0·4864
Huxley	*Brave New World*	0·3675		*Zauberberg*	0·4703
	Antic Hay	0·3777	Jaspers	*Der Philos. Glaube*	0·4968

We see that the entropy values are most characteristic of the different authors and moreover describe quantitatively the order, or binding, relations. In the case of English authors, the entropy values increase in the sequence

Shakespeare (0·29)—Galsworthy (0·32–0·33)—Huxley (0·37–0·38),

according to the transition from the poetic form (higher order) to prose (lower order); in the case of German authors they increase in the sequence

Rilke (0·38)—Hesse (0·47)—Mann (0·47–0·49)—Carossa (0·48)—Jaspers (0·50).

Further, we can make the general statement that authors writing in German are to be found on the whole at higher entropy values than authors writing in English. This means, if we are allowed to generalize from our scarce material, that the German language has, on the whole, a lower order-character than the English language.

Entropy, therefore, fulfills all requirements necessary to a numerical characteristic of style: it is specific, conceptually clear and easy to compute.

(5·4). *A characteristic diagram*

We now plot entropy S against the trace s of the mean-path-matrix (Fig. 1) and get a representation in which every author corresponds to a certain area. The authors of the same language are situated on a certain curve track or curve strip of the (S,s) diagram. This suggests that we get a group of curves or strips with language as a parameter.

6. Summary

The analysis of style, based on the distribution of syllable numbers in the text and on the numerical characteristics deducible from it, yields the result that every work of every author in every language can be associated with a sequence of numbers which characterize certain structural properties of the texts. This characterization is, of course, only possible with respect to the relations of order and coherence of the text elements (syllables), while the 'Sinngehalt' of the text, which is not accessible to such a simple description, will not be covered. Nevertheless, the possibility of a quantitative classification which is very simple to realize has been shown. This yields propositions not only about the peculiarities of style structure of a certain author, but also about the 'carrying structure' of the language on which the individual structure of the author is impressed.

The analogy with respect to the atomistic structure of the solid, liquid and gaseous states which we have repeatedly emphasized for the sake of clearness, is based on the common statistical nature of the underlying states and processes.

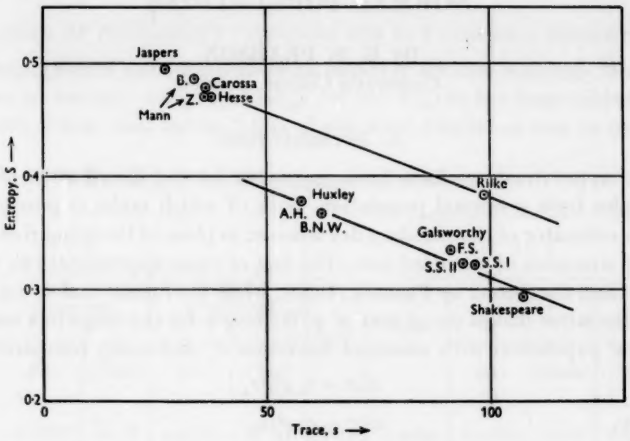

Fig. 1. The characteristic diagram.

The work done on mathematical analysis of style and the collection of material in this respect made by the author go back over a considerable time; the book by Yule (1944) was made known to the author by a footnote in *Nature* (1949, **163**, 688) after most of the material of this paper had been collected.

The author is indebted to his assistant Dipl.-Phys. Wilhelm Frahn for valuable help in condensing his ideas and translating this paper into English and to cand. phys. Herta Weidemann, cand. phys. Helmut Weymann and Emmy Heutz for valuable help in providing the material to establish the frequency distributions.

REFERENCES

BENSE, MAX (1949). *Konturen einer Geistesgeschichte der Mathematik*, **2**. Hamburg: Claassen & Govets.
SHANNON, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379.
YULE, G. U. (1939). On sentence-length as a statistical characteristic of style in prose: with applications to two cases of disputed authorship. *Biometrika*, **30**, 363.
YULE, G. U. (1944). *The Statistical Study of Literary Vocabulary*. Cambridge University Press.

COMPARISON OF TWO APPROXIMATIONS TO THE DISTRIBUTION OF THE RANGE IN SMALL SAMPLES FROM NORMAL POPULATIONS

By E. S. PEARSON
University College, London

1. Introduction

Recently two approximations have been suggested for the distribution of the range in random samples from a normal population, both of which make it possible to use this statistic as an estimator of the standard deviation σ, in place of the usual root-mean-square estimator s, in a number of standard tests. The first of these approximations was published by Cox (1949) and the second by Patnaik (1950); while the former makes use of the distribution of χ^2, the latter makes use of that of χ. Writing w for the range in a sample of size n from a normal population with standard deviation σ, and using comparable notations, Patnaik puts

$$w/\sigma = c_1\chi/\sqrt{\nu_1} \tag{1}$$

while Cox puts

$$w/\sigma = c_2\chi^2/\nu_2, \tag{2}$$

where c_1 and c_2 are scale factors and ν_1 and ν_2 are the 'equivalent' (fractional) degrees of freedom for χ (or χ^2), all four constants being functions of n. The values of these constants are determined so that the first two moments of w/σ and χ (or χ^2) agree.

While the value of ν_2 is obtained readily from the relation $\nu_2 = 2/V_n^2$, where V_n is the coefficient of variation of w in samples of n, the value of ν_1 has to be obtained by inversion of the equation

$$V_n^2 = \tfrac{1}{2}\nu_1\{\Gamma\tfrac{1}{2}\nu_1/\Gamma\tfrac{1}{2}(\nu_1+1)\}^2 - 1. \tag{3}$$

Florin (1950) has given the approximate formula

$$\nu_1 = A^{-1} + 0{\cdot}25 - \tfrac{3}{16}A + \tfrac{3}{64}A^2\ldots, \tag{4}$$

where $A = 2V_n^2$. It will be seen that, as n increases,

$$\nu_2 \sim 4\nu_1 - 1.$$

An implication of the approximation (1) is that w/c_1 may be treated as the usual root-mean-square estimator s, if assigned appropriate degrees of freedom ν_1. These, as one would expect, are always rather less than $n-1$. It follows that w/c_1 may be used in place of s in any test involving a 'Studentized' ratio, the penalty being a reduction in the degrees of freedom with which the error standard deviation is estimated. Cox's approximation does not lend itself to the same procedure since the two sides of equation (2) are of different dimensions, nor can the same interpretation be given to ν_2. However, in the sequential problems with which he was concerned this was not of any consequence.

Both Cox and Patnaik showed that their approximations would be of special value in estimating σ^2 or σ from the mean range $\overline{w}$, in, say, k independent samples of n observations.* If Cox's approximation is used, the mean range will be distributed, apart from a scale factor, as χ^2 with $k\nu_2$ degrees of freedom. Patnaik could not make use of the additive property of

* If the samples are of unequal size, the χ^2-approximation is straightforward; H. A. David (1951, p. 403) has also dealt with the χ-approximation in this case and has given a table (*loc. cit.* p. 409) to assist in computing the weighting factors to be used in combining the sample ranges.

χ^2, but assumed that $\overline{w}/\sigma$ was distributed as $c_1\chi/\sqrt{\nu_1}$, where ν_1 may be obtained approximately from equation (4) with $A = 2V_n^2/k$.

The object of the present note is to compare these two approximations.

2. Comparison of probability integrals for the case of a single sample

The most searching test of the approximations occurs in the case of a single range, i.e. when $k = 1$. The true probability integral of range for $n = 2(1)20$ has been tabled by Pearson & Hartley (1942). From these tables values of w/σ were selected as near as possible to the

Table 1. *Probability integral approximations*

$n = 4$						$n = 6$					
w/σ	True P.I.	χ-approx.		χ^2-approx.		w/σ	True P.I.	χ-approx.		χ^2-approx.	
		P.I.	Error	P.I.	Error			P.I.	Error	P.I.	Error
0·35	0·0053	0·0058	+ 5	0·0012	− 41	0·75	0·0050	0·0060	+10	0·0018	− 32
0·45	0·0111	0·0119	+ 8	0·0036	− 75	0·90	0·0117	0·0134	+17	0·0058	− 59
0·75	0·0483	0·0499	+16	0·0306	−177	1·25	0·0495	0·0526	+31	0·0383	−112
1·00	0·1057	0·1074	+17	0·0877	−180	1·50	0·1031	0·1061	+30	0·0925	−106
1·30	0·2054	0·2065	+11	0·1973	− 81	1·80	0·2000	0·2012	+12	0·1969	− 31
2·00	0·5096	0·5079	−17	0·5303	+207	2·45	0·4899	0·4858	−41	0·5046	+147
2·80	0·8045	0·8031	−14	0·8153	+108	3·25	0·8053	0·8030	−23	0·8119	+ 66
3·25	0·9016	0·9013	− 3	0·9021	+ 5	3·65	0·8981	0·8982	+ 1	0·8978	− 3
3·65	0·9516	0·9520	+ 4	0·9470	− 46	4·05	0·9519	0·9533	+14	0·9481	− 38
4·40	0·9899	0·9904	+ 5	0·9849	− 50	4·75	0·9898	0·9910	+12	0·9862	− 36
4·70	0·9951	0·9955	+ 4	0·9911	− 40	5·05	0·9952	0·9960	+ 8	0·9925	− 27
$n = 10$						$n = 15$					
w/σ	True P.I.	χ-approx.		χ^2-approx.		w/σ	True P.I.	χ-approx.		χ^2-approx.	
		P.I.	Error	P.I.	Error			P.I.	Error	P.I.	Error
1·35	0·0054	0·0076	+22	0·0034	−20	1·80	0·0049	0·0077	+ 28	0·0041	− 8
1·50	0·0117	0·0149	+32	0·0085	−32	1·95	0·0108	0·0150	+ 42	0·0096	−12
1·85	0·0479	0·0530	+51	0·0425	−54	2·30	0·0468	0·0535	+ 67	0·0451	−17
2·10	0·1015	0·1061	+46	0·0971	−44	2·55	0·1026	0·1083	+ 57	0·1015	−11
2·40	0·2025	0·2034	+ 9	0·2022	− 3	2·85	0·2103	0·2105	+ 2	0·2109	+ 6
3·00	0·4878	0·4800	−78	0·4954	+76	3·40	0·4885	0·4775	−110	0·4909	+24
3·75	0·8062	0·8026	−36	0·8089	+27	4·10	0·8036	0·7990	− 46	0·8043	+ 7
4·15	0·9038	0·9047	+ 9	0·9030	− 8	4·45	0·8964	0·8975	+ 11	0·8962	− 2
4·45	0·9474	0·9500	+26	0·9453	−21	4·80	0·9505	0·9543	+ 38	0·9498	− 7
5·15	0·9898	0·9919	+21	0·9880	−18	5·45	0·9900	0·9928	+ 28	0·9895	− 5
5·40	0·9948	0·9962	+14	0·9934	−14	5·70	0·9950	0·9968	+ 18	0·9946	− 4

median and the lower and upper 20, 10, 5, 1 and 0·5 % points for samples with $n = 4$, 6, 10 and 15. These 11 values of w/σ and the corresponding true probability integral (P.I.) are shown in Table 1 for the four values of n. Using the appropriate scale factors c_1, c_2 and degrees of freedom ν_1, ν_2,* it was then possible to compute the corresponding probability integrals for both the χ and the χ^2 approximations, with the help of the *Tables of the Incomplete Gamma Function* (K. Pearson, 1922). In certain cases, near the start of the curve where interpolation in the Tables is difficult, the probability integral was calculated from a few terms of the series

$$P(\chi^2 \mid \nu) = e^{-m} \sum_{j=0}^{\infty} m^{d+j}/\Gamma(d+j+1),$$

where $m = \frac{1}{2}\chi^2, d = \frac{1}{2}\nu$. The resulting approximate probability integrals are shown in Table 1 as well as the differences

$$\text{Error} = \text{Approximate P.I.} - \text{True P.I.}$$

It will be seen that, up to $n = 15$:

(*a*) The errors are of opposite sign, the χ-approximation having a distribution less skew and the χ^2-approximation a distribution more skew than the true distribution.

(*b*) In both cases, the error is greater at the lower than at the upper tail of the curves.

(*c*) Broadly speaking, the approximations are of equal accuracy at $n = 10$. For $n < 10$ the χ-approximation and for $n > 10$ the χ^2-approximation is the more accurate.

No simple criterion can be given to discriminate between an approximation which is adequate and one which is not. What may be quite satisfactory in a simple test of significance may lead to more serious error in the distribution of derived statistics. In § 4 below, mention is made of the possible use of the ratio of maximum to minimum range in a set of k independent ranges as a test for heterogeneity of variance. Here, the form of the range distribution in the tails becomes of increasing importance as k increases, and before the test referred to is based on either of these approximations, some investigation into the effect of the 'errors' given in Table 1 is needed.

3. The β_1, β_2 diagram

Both approximate distributions have the correct mean and variance, whether a single range or a mean range is used. Their accuracy will, therefore, depend on the extent to which they have the correct 'shape'. This may be usefully examined by calculating the moment ratios

$$\beta_1 = \gamma_1^2 = \mu_3^2/\mu_2^3 \quad \text{and} \quad \beta_2 = \gamma_2 + 3 = \mu_4/\mu_2^2,$$

and plotting the resulting β_1, β_2 points for the true and approximate distributions. Four main series of points are shown in Fig. 1:

(*a*) Those for the distribution of χ (or s). These points lie on a curve which sweeps round to the Normal point (0, 3) as the degrees of freedom ν increase. The values of $\beta_1(s)$ and $\beta_2(s)$ were taken from *Tables for Statisticians and Biometricians*, Part II, Table XVII.

(*b*) Those for the distribution of χ^2 (or s^2). These lie on the type III line $2\beta_2 - 3\beta_1 - 6 = 0$, and converge on the Normal point as ν increases much more slowly than the corresponding points for χ.

(*c*) Those for the distribution of the range w. It is the peculiar character of this curve,

* Numerical values of the constants have been given by Cox (1949), Patnaik (1950) and (for c_1, ν_1) by David (1951). As greater accuracy was needed, however, values were recomputed from the appropriate formulae.

which first swings in and then turns away from the Normal point as n increases, that makes approximation by a single type of curve difficult. The values of $\beta_1(w)$ and $\beta_2(w)$ were taken from the table given by Hartley & Pearson (1951).

(*d*) Those for the distribution of $\log_e \chi^2$ (or $\log_e s^2$). The β_1, β_2 values were obtained from the asymptotic expressions given by Bartlett & Kendall (1946), which are correct to the accuracy needed for $\nu \geqslant 20$. While this distribution is not strictly relevant to the present inquiry, it seemed of interest to include a plot of the points.* It will be noticed that while, for a given ν, the distribution of $\log_e s^2$ is more nearly normal than that of s^2, it is considerably further from the normal than the distribution of s. The $\log_e s^2$ distribution is negatively skew, i.e. $\gamma_1 = \sqrt{\beta_1} < 0$, and in this sense the positive skewness of the s^2 distribution may be said to have been overcorrected by the transformation. The great merit of the logarithmic transformation is, of course, that it gives a statistic whose variance is independent of σ^2.

Without further aid we cannot interpret a diagram like Fig. 1, since 'distances' in different directions and in different parts of the field are not in constant proportion to differences in probability integrals. Experience has, however, shown that the great majority of sampling distributions commonly met are well represented by curves of the Pearson system having the same β_1, β_2 values. This means that Fig. 1 may be interpreted with the help of Pearson & Merrington's (1951) tables of the upper and lower 5 and 0·5 % of this system, expressed in standard measure.† A brief extract from these tables is given in Table 2 below. This shows that within the region of the field with which we are concerned, changes in β_1 are of much more importance than changes in β_2.

Table 2. *Standardized deviates,* $(x-\mu)/\sigma$ *for curves of the Pearson system.* (*Positive skewness, i.e.* $\mu_3 > 0$)

β_2 \ β_1	0·00	0·05	0·15	0·30	0·00	0·05	0·15	0·30
	Lower 5 % points				Upper 5 % points			
3·0	1·64	1·58	1·52	1·44	1·64	1·71	1·76	1·82
3·2	1·64	1·58	1·53	1·46	1·64	1·70	1·75	1·80
3·4	1·64	1·58	1·53	1·47	1·64	1·69	1·74	1·79
	Lower 0·5 % points				Upper 0·5 % points			
3·0	2·58	2·33	2·11	1·84	2·58	2·76	2·86	2·93
3·2	2·65	2·42	2·20	1·95	2·65	2·83	2·93	3·01
3·4	2·71	2·48	2·28	2·04	2·71	2·88	2·99	3·07

* These points lie very close to the Type V line of K. Pearson's system and also not far from the line representing log-normal distributions, but it is not clear that this correspondence can be turned to any useful account.

† For example, to the accuracy with which interpolation in these tables is possible, we obtain from them correct values of the 5 and 0·5 % points of the s and w distributions represented in Fig. 1 except for small errors in the lower 0·5 % points for s when $\nu = 3$ and for w when $n = 4$.

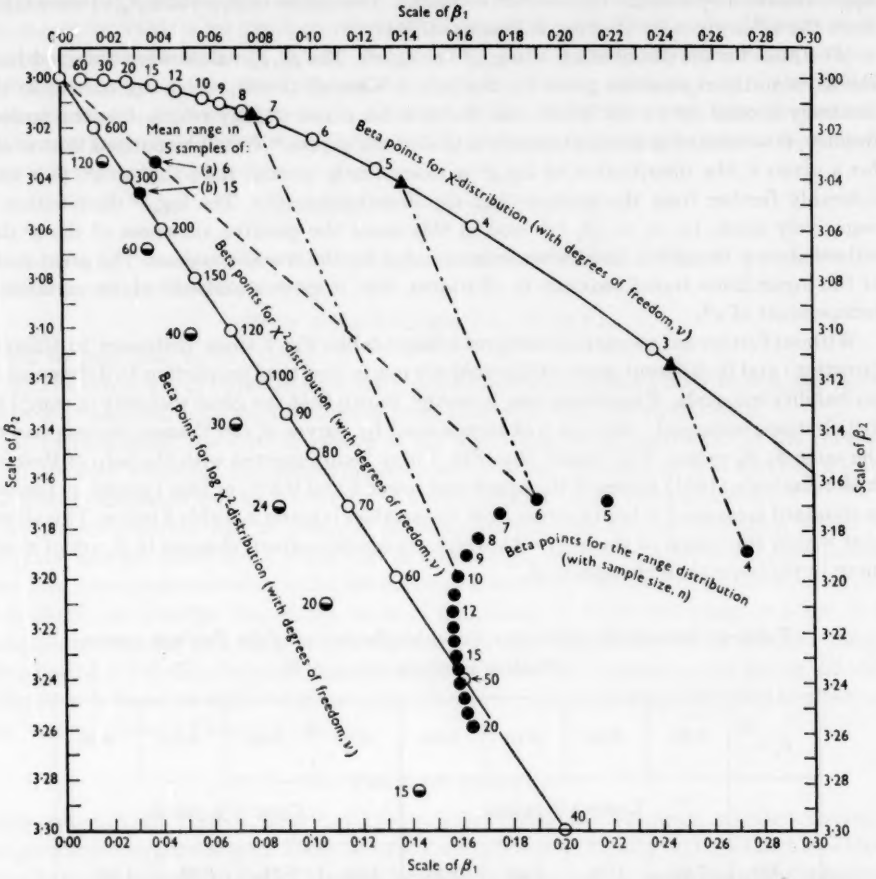
Scale of β_1
Scale of β_2
Mean range in 5 samples of:
(a) 6
(b) 15
Beta points for χ-distribution (with degrees of freedom, ν)
Beta points for χ^2-distribution (with degrees of freedom, ν)
Beta points for log χ^2-distribution (with degrees of freedom, ν)
Beta points for the range distribution (with sample size, n)

Fig. 1. Diagram showing relation between β_1, β_2 points of various sampling distributions.

Table 3. *Comparison of approximations to the distribution of the range*

n	χ-approx. ν_1	χ^2-approx.		
		ν_2	β_1	β_2
4	2·93	10·95	0·731	4·10
6	4·68	17·86	0·448	3·67
10	7·68	29·82	0·268	3·40
15	10·77	42·16	0·189	3·28

Consider now the distribution of range in a single sample. As an illustration, Table 3 gives for $n = 4, 6, 10$ and 15 the degrees of freedom ν_1 and ν_2 of the two approximations and the β_1, β_2 values of the χ^2-approximation (as the first three of these fall outside Fig. 1). For $n = 4, 6$ and 10 the β_1, β_2 points of the χ-approximation are marked by triangles in Fig. 1; these have been joined by broken lines to the corresponding true points for range. The effect of the divergence of the beta-curves for w and χ is clear. After good agreement when $n = 4$, there is a gradual deterioration in fit as n increases, but this does not become appreciable until n reaches, say, 7 or 8. Even at $n = 10$, Table 1 shows fair agreement in the probability integrals.

For the χ^2-approximation, the β_1, β_2 points are far outside the diagram for $n = 4$ and 6, so that the distribution is much too skew. At $n = 10$ the points for the two approximations are at about equal distances away from the true point, but in opposite directions. When $n = 15$ the true and χ^2-approximation points are still approaching one another; at $n = 20$, with $\nu_2 = 52{\cdot}54$, they have passed and at greater sample sizes will continue to diverge.

We may next consider very briefly the relation of the approximations to the distribution of $\overline{w}$, the mean of k ranges. The true β_1, β_2 point of $\overline{w}$ will lie on the straight line joining the Normal point (0, 3) to the point $\beta_1(w), \beta_2(w)$, such that $\beta_1(\overline{w}) = \beta_1(w)/k$. Fig. 1 shows the points for the mean range in $k = 5$ samples of (*a*) $n = 6$ and (*b*) $n = 15$ observations. For these cases the equivalent degrees of freedom of the two approximations are:

	ν_1 for χ-approx.	ν_2 for χ^2-approx.
5 samples of 6	22·6	89·3
5 samples of 15	52·9	210·8

Although no detailed calculations have been made, it appears that the χ-approximation will be the more accurate for $n = 6$ and the χ^2-approximation for $n = 15$. As k increases, all distributions for mean range are, of course, becoming nearer and nearer to the normal, and any differences will become progressively of less account.

4. A TEST FOR HETEROGENEITY OF VARIANCE

Hartley (1950) has suggested that if a number k of independent variance estimates $s_t^2 (t = 1, 2, \ldots, k)$ are available, each based on the same number of degrees of freedom ν, then a short-cut test of the hypothesis that the variances in the k sampled populations (assumed normal) are the same, may be made using the ratio $s^2(\text{max.})/s^2(\text{min.})$ of the maximum to the minimum values of the k estimates. He provided an approximate table of the upper 5 % points of this ratio for $k = 2(1)\,12$ and $\nu = 2(1)\,10, 12, 15, 20, 30, 60, \infty$. It is hoped to publish shortly an accurate table of both 5 % and 1 % points of $s^2(\text{max.})/s^2(\text{min.})$.*

Clearly, to the accuracy involved in the approximations, we may use an even quicker test of variance heterogeneity by noting the maximum and minimum ranges among the k groups of observations, finding the ratio $w(\text{max.})/w(\text{min.})$ and then using Hartley's table:

(*a*) with ν_1 degrees of freedom and the square of the range ratio for the χ-approximation;

(*b*) with ν_2 degrees of freedom and the range ratio itself for the χ^2-approximation.

No calculation of the scale factors c_1 or c_2 is required. The method needs further investigation, but it might be supposed that more accurate results would be obtained with the χ-approximation when $n < 10$ and with the χ^2-approximation when $n > 10$.

* This table has now been computed by Mr H. A. David.

Illustration

Suppose that we have 5 samples of 10 observations. What is the 5 % level for the ratio w(max.)/w(min.) if the samples have been drawn independently from normal populations having a common variance?

(*a*) *χ-approximation.* $\nu_1 = 7{\cdot}68$. Interpolating in Hartley's table (as adjusted by David), it is found that for $k = 5$, $\nu = 7{\cdot}68$ the 5 % limit for s^2(max.)/s^2(min.) is 8·59. The square root of this value, or 2·93, gives the approximation to the 5 % limit for the range ratio.

(*b*) *χ^2-approximation.* $\nu_2 = 29{\cdot}82$. Interpolation in Hartley's adjusted table gives, directly, 2·79 as the 5 % point for the ratio.

The difference between 2·93 and 2·79 no doubt lies in the fact, brought out in Table 1, that the probability integral at the tails of the χ and χ^2-approximations differ in opposite senses from the true probability integral. At a guess, the true 5 % point for w(max.)/w(min.) when $k = 5$, $n = 10$, may lie midway between the two approximations, at about 2·86.

While a test based on w(max.)/w(min.) will in general be less powerful than the well-known test for heterogeneity of variance (Neyman & Pearson (1931), Bartlett (1937)), there is clearly much to be said for a short-cut test based on this range ratio. If tables are available, its application in a preliminary survey of data is immediate. It also conforms with the intuitive method of judging whether there is heterogeneity by comparing the range of observations in the group having greatest spread with that in the group having least spread.

REFERENCES

Bartlett, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.
Bartlett, M. S. & Kendall, D. G. (1946). *J.R. Statist. Soc. Suppl.* **8**, 128.
Cox, D. R. (1949). *J.R. Statist. Soc.* B, **11**, 101.
David, H. A. (1951). *Biometrika*, **38**, 393.
Florin, H. (1950). *Commun. Roy. Flem. Acad.* (Science Series), **12**, 6.
Hartley, H. O. (1950). *Biometrika*, **37**, 308.
Hartley, H. O. & Pearson, E. S. (1951). *Biometrika*, **38**, 463.
Neyman, J. & Pearson, E. S. (1931). *Bull. Int. Acad. Cracovie*, A, 460.
Patnaik, P. B. (1950). *Biometrika*, **37**, 78.
Pearson, E. S. & Hartley, H. O. (1942). *Biometrika*, **32**, 302.
Pearson, E. S. & Merrington, Maxine (1951). *Biometrika*, **38**, 4.
Pearson, K. (1922). *Tables of the Incomplete Gamma Function.* Cambridge University Press.

THE COVERING CIRCLE OF A SAMPLE FROM A CIRCULAR NORMAL DISTRIBUTION

By H. E. DANIELS

Statistical Laboratory, University of Cambridge

1. Introduction

Consider a sample of n observations from the circular normal distribution

$$dF = e^{-(x^2+y^2)/2\sigma^2}\frac{dx\,dy}{2\pi\sigma^2}.$$

The *covering circle* of the sample is defined to be the smallest circle in the x, y plane containing on or within it every sample point. The purpose of this paper is to study the distribution of the radius and centre of the covering circle. They are bivariate analogues of the half-range and mid-range point of a univariate normal sample.

My attention was drawn to the problem by Mr D. Fraser of the Psychological Laboratory, Cambridge, who used the radius of the covering circle as a convenient rapid measure of dispersion.

2. Some preliminary formulae

We shall require some known results concerning a simpler problem of the same kind (see, for example, Quenouille, 1949).

The chance that n sample points lie within a circle of radius r centred at a given distance ρ from the true mean is $P^n(r,\rho)$, where*

$$P(r,\rho) = e^{-\rho^2/2\sigma^2}\int_0^r \frac{s\,ds}{\sigma^2}e^{-s^2/2\sigma^2}I_0\left(\frac{s\rho}{\sigma^2}\right), \tag{2·1}$$

$I_\nu(z)$ being the Bessel function of imaginary argument and order ν, which is given by

$$I_\nu(z) = \frac{1}{2\pi}\int_0^{2\pi} e^{z\cos\theta}\cos\nu\theta\,d\theta,$$

when ν is an integer. Since $I_0'(z) = I_1(z)$,

$$\frac{\partial P(r,\rho)}{\partial\rho} = -\frac{\rho}{\sigma^2}P(r,\rho) + e^{-\rho^2/2\sigma^2}\int_0^r \frac{s^2\,ds}{\sigma^4}e^{-s^2/2\sigma^2}I_1\left(\frac{s\rho}{\sigma^2}\right),$$

whence, integrating by parts, using the fact that

$$\frac{d}{dz}\{zI_1(z)\} = I_0(z),$$

we find

$$\frac{\partial P(r,\rho)}{\partial\rho} = -\frac{r}{\sigma^2}e^{-(r^2+\rho^2)/2\sigma^2}I_1\left(\frac{r\rho}{\sigma^2}\right), \tag{2·2}$$

and

$$P(r,\rho) = \frac{r}{\sigma^2}e^{-r^2/2\sigma^2}\int_\rho^\infty e^{-u^2/2\sigma^2}I_1\left(\frac{ru}{\sigma^2}\right)du. \tag{2·3}$$

* This function has been discussed in classified reports by W. R. Hynd and the author in Britain and A. N. Lowan in the U.S.A. The function $1-P$ has been extensively tabulated by the Rand Corporation.

Also
$$P(r,\rho) = 1 - e^{-\rho^2/2\sigma^2}\int_r^\infty \frac{s\,ds}{\sigma^2} e^{-s^2/2\sigma^2} I_0\left(\frac{s\rho}{\sigma^2}\right)$$
$$= 1 - e^{-(r^2+\rho^2)/2\sigma^2} I_0\left(\frac{r\rho}{\sigma^2}\right) - P(\rho,r). \tag{2·4}$$

Further partial integration gives the expansion

$$1 - P(r,\rho) = e^{-(r^2+\rho^2)/2\sigma^2}\left\{I_0\left(\frac{r\rho}{\sigma^2}\right) + \frac{\rho}{r} I_1\left(\frac{r\rho}{\sigma^2}\right) + \frac{\rho^2}{r^2} I_2\left(\frac{r\rho}{\sigma^2}\right) + \ldots\right\}, \tag{2·5}$$

and similarly, from (2·1),

$$P(r,\rho) = e^{-(r^2+\rho^2)/2\sigma^2}\left\{\frac{r}{\rho} I_1\left(\frac{r\rho}{\sigma^2}\right) + \frac{r^2}{\rho^2} I_2\left(\frac{r\rho}{\sigma^2}\right) + \ldots\right\}. \tag{2·6}$$

Both series converge for all r and ρ, their sum yielding the familiar Laurent development

$$e^{(r^2+\rho^2)/\sigma^2} = \sum_{m=-\infty}^{\infty} \left(\frac{r}{\rho}\right)^m I_m\left(\frac{r\rho}{\sigma^2}\right).$$

3. Joint distribution of the radius and centre of the covering circle

Samples may be divided into two main categories, for which the covering circle has respectively the following properties:

(i) The circle passes through three of the points forming an acute-angled triangle, and the remainder of the points lie inside the circle.

(ii) The circle passes through two diametrically opposite points, the remainder lying inside the circle.

These two contingencies are mutually exclusive, and the remaining cases where the circle passes through more than three points can be shown to be relatively of negligible probability. The required joint distribution is therefore obtained by adding the probabilities for cases (i) and (ii), which we now evaluate.

For samples of type (i) the chance that the radius lies between r and $r+dr$ and that the centre is at distance ρ to $\rho+d\rho$ from the origin is

$$dF_1 = \tfrac{1}{3}n(n-1)(n-2)\,dQ_1 P^{n-3}(r,\rho), \tag{3·1}$$

where dQ_1 is the chance that three 'labelled' points lie in a specified order round the circle, and $P(r,\rho)$ is given by (2·1). The factor $n(n-1)(n-2)$ is the number of ways of selecting the three points, and the divisor 3 accounts for the identity of orders 123, 231, 312 on the circle. To find dQ_1, let the three points have coordinates $x_j, y_j\,(j = 1, 2, 3)$ and write

$$x_j = \rho\cos\phi + r\cos\theta_j, \quad y_j = \rho\sin\phi + r\sin\theta_j,$$

where ρ, ϕ are the polar co-ordinates of the centre, and $0 \leqslant \theta_1 \leqslant \theta_2 \leqslant \theta_3 < 2\pi$. The Jacobian is

$$\frac{\partial(x_1,y_1,x_2,y_2,x_3,y_3)}{\partial(\rho,r,\phi,\theta_1,\theta_2,\theta_3)} = r^3\rho\{\sin(\theta_2-\theta_1) + \sin(\theta_3-\theta_2) + \sin(\theta_1-\theta_3)\}, \tag{3·2}$$

and the points will always form an acute-angled triangle if

$$\left.\begin{aligned} 0 &\leqslant \theta_1 < 2\pi, \\ \theta_1 &\leqslant \theta_2 < \theta_1 + \pi, \\ \theta_1 + \pi &\leqslant \theta_3 < \theta_2 + \pi. \end{aligned}\right\} \tag{3·3}$$

The probability element for the three points transforms to

$$\exp\left[\frac{-3(r^2+\rho^2)}{2\sigma^2}\right]\frac{r^3\rho\,dr\,d\rho}{(2\pi\sigma^2)^3}\exp\left[-\frac{\rho r}{\sigma^2}\{\cos(\theta_1-\phi)+\cos(\theta_2-\phi)+\cos(\theta_3-\phi)\}\right]$$
$$\times\{\sin(\theta_2-\theta_1)+\sin(\theta_3-\theta_2)+\sin(\theta_1-\theta_3)\}\,d\phi\,d\theta_1\,d\theta_2\,d\theta_3,$$

and dQ_1 is obtained on integrating the θ's over (3·3) and ϕ from 0 to 2π.

Since, by symmetry, integration of the θ's must give a result independent of ϕ, the ϕ integration merely introduces a factor 2π. The θ integrations can be effected by putting $\phi = 0$, expanding $\sin(\theta_2-\theta_1)+\sin(\theta_3-\theta_2)+\sin(\theta_1-\theta_3)$ and integrating the six terms separately. After some manipulation it is found that

$$dQ_1 = \frac{3}{\sigma^4}\exp\left[-\frac{3(r^2+\rho^2)}{2\sigma^2}\right]I_1\left(\frac{r\rho}{\sigma^2}\right)r^2\,dr\,d\rho. \tag{3·4}$$

For samples of type (ii) the chance of r to $r+dr$, ρ to $\rho+d\rho$, is by a similar argument

$$dF_2 = \tfrac{1}{2}n(n-1)\,dQ_2\,P^{n-2}(r,\rho), \tag{3·5}$$

where dQ_2 is the chance that two labelled points are at diametrically opposite positions on the circle.

The two points may be given co-ordinates

$$x_1 = \rho\cos\phi + r\cos\theta, \quad y_1 = \rho\sin\phi + r\sin\theta,$$
$$x_2 = \rho\cos\phi - r\cos\theta, \quad y_2 = \rho\sin\phi - r\sin\theta,$$

where $0 \leqslant \theta < 2\pi$, $0 \leqslant \phi < 2\pi$, and since

$$\frac{\partial(x_1,y_1,x_2,y_2)}{\partial(\rho,r,\phi,\theta)} = 4r\rho,$$

it follows without difficulty that

$$dQ_2 = \frac{4}{\sigma^4}e^{-(r^2+\rho^2)/\sigma^2}r\rho\,dr\,d\rho. \tag{3·6}$$

The joint distribution of r and ρ is therefore

$$dF_n(r,\rho) = dF_1 + dF_2 = \left\{\frac{n(n-1)(n-2)}{\sigma^4}\exp\left[-\frac{3}{2\sigma^2}(r^2+\rho^2)\right]r^2I_1\left(\frac{r\rho}{\sigma^2}\right)P^{n-3}(r,\rho)\right.$$
$$\left.+\frac{2n(n-1)}{\sigma^4}\exp\left[-\frac{(r^2+\rho^2)}{\sigma^2}\right]r\rho\,P^{n-2}(r,\rho)\right\}dr\,d\rho,$$

which reduces, by (2·2), to the form

$$dF_n(r,\rho) = -n(n-1)\,e^{-r^2/\sigma^2}\frac{r\,dr}{\sigma^2}\frac{\partial}{\partial\rho}\{e^{-\rho^2/\sigma^2}P^{n-2}(r,\rho)\}\,d\rho. \tag{3·7}$$

4. Distribution of the radius

On integrating ρ from 0 to ∞ the distribution of r is found to be

$$dF_n(r) = n(n-1)\,e^{-r^2/\sigma^2}(1-e^{-r^2/2\sigma^2})^{n-2}\frac{r\,dr}{\sigma^2}. \tag{4·1}$$

This may be contrasted with

$$d\{P^n(r,0)\} = n\,e^{-r^2/2\sigma^2}(1-e^{-r^2/2\sigma^2})^{n-1}\frac{r\,dr}{\sigma^2},$$

which is the distribution of the radius of the smallest circle, centred at the true mean, enclosing all sample points. On the other hand, it is remarkable that (4·1) is the same as the distribution of the radius of the smallest circle, centred at the true mean, enclosing *all but one* of the sample points.

The distribution function for r can be written

$$F_n(r) = n(1-e^{-r^2/2\sigma^2})^{n-1} - (n-1)(1-e^{-r^2/2\sigma^2})^n, \tag{4·2}$$

or it may be computed from tables of the Incomplete B-function (Pearson, 1934) as $I_z(2, n-1)$, with $z = e^{-r^2/2\sigma^2}$ in the notation of the tables quoted. Upper and lower confidence limits for σ, given r, can also be found from the relation

$$\frac{r}{\sigma} = (2\log\{\tfrac{1}{2}(n-1)\mathbf{F}^{-1}+1\})^{\frac{1}{2}}, \tag{4·3}$$

where $\mathbf{F}$ has Fisher's variance ratio distribution with 4 and $2(n-1)$ degrees of freedom (see Table 1).

Table 1. *Distribution of r/σ. Mean, variance and percentage points*

n	Mean	Variance	1%	5%	95%	99%
2	0·886	0·2146	0·100	0·226	1·731	2·146
3	1·211	0·1990	0·348	0·539	2·000	2·380
4	1·409	0·1827	0·541	0·756	2·157	2·518
5	1·548	0·1700	0·709	0·916	2·268	2·616
6	1·655	0·1601	0·835	1·041	2·352	2·691
7	1·742	0·1522	0·939	1·142	2·421	2·752
8	1·814	0·1456	1·027	1·228	2·478	2·803
9	1·876	0·1401	1·103	1·301	2·527	2·847
10	1·929	0·1354	1·170	1·365	2·570	2·886
11	1·977	0·1314	1·229	1·421	2·608	2·920
12	2·020	0·1278	1·282	1·472	2·642	2·951
13	2·058	0·1245	1·330	1·517	2·673	2·979
14	2·093	0·1214	1·374	1·559	2·702	3·005
15	2·125	0·1183	1·414	1·597	2·728	3·028
20	2·255	0·1095	1·576	1·750	2·833	3·124
30	2·427	0·0977	1·790	1·953	2·975	3·254
40	2·543	0·0906	1·932	2·087	3·071	3·342
50	2·629	0·0855	2·037	2·187	3·143	3·408
100	2·881	0·0731	2·341	2·477	3·358	3·608

The moment generating function for r^2 is

$$E(e^{r^2 t}) = \frac{\Gamma(n+1)\,\Gamma(2-2t\sigma^2)}{\Gamma(n+1-2t\sigma^2)} = \prod_{m=2}^{n} \frac{1}{(1-2t\sigma^2/m)}, \tag{4·4}$$

and the even moments of r about the origin, $\mu'_{2s} = E(r^{2s})$, are obtained by the usual relations from the cumulants of r^2,

$$\kappa_j(r^2) = (j-1)!\,2^j\sigma^{2j} \sum_{m=2}^{n} \frac{1}{m^j}. \tag{4·5}$$

In particular
$$\mu_2' = \kappa_1(r^2) = 2\sigma^2\left(\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n}\right)$$
$$\sim 2\sigma^2(\log n - 1 + \gamma) \tag{4·6}$$

for large n, where γ is Euler's constant 0·577216....

The following two formulae are useful for moments of any order. From (4·1),

$$dF_n(r) = \sum_{m=0}^{n}\binom{n}{m}(-)^m m(m-1)\,e^{-mr^2/2\sigma^2}\frac{r\,dr}{\sigma^2}$$
$$= (-)^n\,\Delta^n\{m(m-1)\,e^{-mr^2/2\sigma^2}\}_{m=0}\frac{r\,dr}{\sigma^2},$$

Δ operating on m. Hence

$$\mu_s' = E(r^s) = (-)^n\,2^{\frac{1}{2}s}\Gamma(\tfrac{1}{2}s+1)\,\sigma^s\Delta^n y_0, \tag{4·7}$$

where $y_0 = 0$, $y_1 = 0$, $y_m = (m-1)/m^{\frac{1}{2}s}$ $(m \geqslant 2)$. This formula is convenient for calculating μ_s'/σ^s for small n, but becomes inaccurate when n exceeds about 10.

For large n there is the asymptotic expansion

$$\mu_s' = E\{r^2 - \mu_2' + \mu_2'\}^{\frac{1}{2}s}$$
$$= \mu_2'^{\frac{1}{2}s}\sum_m \frac{\frac{1}{2}s(\frac{1}{2}s-1)\ldots(\frac{1}{2}s-m+1)}{m!}\frac{\mu_m(r^2)}{\mu_2'^m}, \tag{4·8}$$

where $\mu_m(r^2)$ is the mth moment of r^2 about its mean, calculated from the known cumulants (4·5) in the ordinary way. But since $\mu_2' \sim O(\log n)$, n has to be fairly large for reasonable accuracy in calculating the odd moments.

Table 1 gives the mean and variance of r/σ for a range of values of n. It was computed by using (4·7) and (4·8) at each end of the range and by numerical integration for the intermediate values. Note that $2r$ is always greater than the corresponding mean range for a univariate normal sample of n, as is otherwise evident from the fact that the diameter of the covering circle is the greatest of the ranges of all univariate projections of the sample, and hence its expectation must be greater than that of the mean range of all such projections.

When n is large, (4·8) and (4·5) give

$$E(r) = \mu_1' \sim \mu_2'^{\frac{1}{2}} \sim \sigma\sqrt{(2\log n)},$$
$$\operatorname{var} r = \mu_2' - \mu_1'^2 \sim \frac{1}{4}\frac{\kappa_2(r^2)}{\mu_2'} \sim \left(\frac{\pi^2}{12}-\frac{1}{2}\right)\frac{\sigma^2}{\log n}.$$

The effective range of r is therefore such that $e^{-r^2/2\sigma^2} = O(1/n)$, and (4·1) has the limiting form

$$dF_n(r) \sim n^2\exp\left[-\frac{r^2}{\sigma^2} - n\,e^{-r^2/2\sigma^2}\right]\frac{r\,dr}{\sigma^2}\left\{1+O\left(\frac{1}{n}\right)\right\}, \tag{4·9}$$

so that $n\,e^{-r^2/2\sigma^2}$ is approximately χ^2 with 4 degrees of freedom. For very large n, (4·9) in turn approximates to the general form given by Gumbel (1936) for a quantile, but the neglected terms are then $O(1/\log n)$.

5. Distribution of the centre

The distribution function for the distance ρ of the centre from the true mean is, from (3·7),

$$F_n(\rho) = 1 - n(n-1)\, e^{-\rho^2/\sigma^2} \int_0^\infty e^{-r^2/\sigma^2} P^{n-2}(r,\rho) \frac{r\,dr}{\sigma^2}. \tag{5·1}$$

This does not appear to reduce further in the general case. For $n = 3$, however, it takes the following simple form:

$$\begin{aligned} F_3(\rho) &= 1 - 6\, e^{-\rho^2/\sigma^2} \int_0^\infty e^{-r^2/\sigma^2} \frac{r\,dr}{\sigma^2} \int_0^r \frac{s\,ds}{\sigma^2} e^{-s^2/2\sigma^2} I_0\left(\frac{s\rho}{\sigma^2}\right) \\ &= 1 - 3\, e^{-3\rho^2/2\sigma^2} \int_0^\infty e^{-3r^2/2\sigma^2} I_0\left(\frac{r\rho}{\sigma^2}\right) \frac{r\,dr}{\sigma^2} \text{ (by parts)} \\ &= 1 - e^{-4\rho^2/3\sigma^2}, \end{aligned} \tag{5·2}$$

which may be compared with the distribution function $1 - e^{-3\rho^2/2\sigma^2}$ for the centroid of the three points.

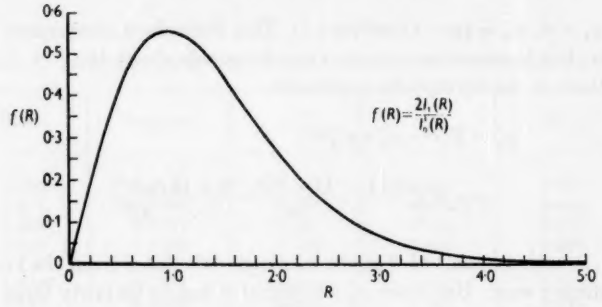

Fig. 1.

When n is large it was shown that $r \sim O(\sqrt{\log n})$. It will appear that if we assume also that $\rho \sim O(1/\sqrt{\log n})$, a limiting distribution function is obtained which ranges from 0 to 1 so that no other values of ρ need be considered. With r and ρ of the orders stated, we have from (2·5),

$$\begin{aligned} P(r,\rho) &\sim 1 - e^{-(r^2+\rho^2)/2\sigma^2} I_0\left(\frac{r\rho}{\sigma^2}\right) + O\left(\frac{\rho}{r} e^{-r^2/2\sigma^2}\right) \\ &\sim 1 - e^{-r^2/2\sigma^2} I_0\left(\frac{r\rho}{\sigma^2}\right) + O\left(\frac{1}{n \log n}\right), \end{aligned}$$

and so (5·1) becomes

$$F_n(\rho) \sim 1 - n^2 \int_0^\infty \frac{r\,dr}{\sigma^2} \exp\left[-\frac{r^2}{\sigma^2} - n\, e^{-r^2/2\sigma^2} I_0\left(\frac{r\rho}{\sigma^2}\right)\right] \left\{1 + O\left(\frac{1}{\log n}\right)\right\}.$$

Write

$$w = n\, e^{-r^2/2\sigma^2}, \quad R = \frac{\rho}{\sigma}\sqrt{(2 \log n)}.$$

Then

$$\begin{aligned} \frac{r\rho}{\sigma^2} &= R\left(1 - \frac{\log w}{\log n}\right)^{\frac{1}{2}}, \\ I_0\left(\frac{r\rho}{\sigma^2}\right) &= I_0(R) - I_1(R) \frac{R \log w}{2 \log n} + \ldots, \\ e^{-w I_0(r\rho/\sigma^2)} &= e^{-w I_0(R)} \left\{1 + I_1(R) \frac{Rw \log w}{2 \log n} + \ldots\right\}. \end{aligned}$$

Hence
$$F_n(\rho) \sim 1 - \int_0^\infty w e^{-w I_0(R)}\, dw + O\left(\frac{1}{\log n}\right)$$
$$= 1 - \frac{1}{I_0^2(\rho/\sigma\sqrt{(2\log n)})} + O\left(\frac{1}{\log n}\right), \tag{5·3}$$

which is the limiting distribution function. The frequency function for R is

$$f(R) = \frac{2I_1(R)}{I_0^3(R)}.$$

Fig. 1 shows the form of $f(R)$, which behaves like R for small R, and $4\pi R^2 e^{-R}$ when R is large.

I am indebted to Mr D. A. East for the calculations of Table 1.

REFERENCES

GUMBEL, E. J. (1936). *Ann. Inst. Poincaré*, **5**, 115.
PEARSON, K. (1934). *Tables of the Incomplete* B-*function*. London: Biometrika Office.
QUENOUILLE, M. H. (1949). *Proc. Edinb. Math. Soc.*, ser. 2, **8**, 95.

THE FREQUENCY JUSTIFICATION OF CERTAIN SEQUENTIAL TESTS

By G. A. BARNARD, *Imperial College, London*

1. The primary object of the following paper is to give a 'frequency justification', in the sense of Neyman and Pearson, of a form of sequential t-test proposed by Barnard (1949). When we have done this, we go on to show that the result can be generalized, so that any sequential test arrived at by the argument used in the earlier paper will have a 'frequency justification'. Finally, it is pointed out that the discussion has a bearing on the problem of 'reconciling' the three main theories of statistical inference, represented nowadays by Professors Fisher, Jeffreys, and Neyman and Pearson.

2. To explain the meanings of the words we have put in inverted commas, we will consider first the case of testing two simple hypotheses $\mathscr{H}_j\,(j = 1, 2)$ against each other. If E_n represents the result of n independent trials, to say that the hypotheses are simple means that $\Pr\{E_n \mid \mathscr{H}_j\}$ can be calculated as a definite number for each possible result and for each of the two hypotheses. For a fixed hypothesis, for varying results, $\Pr\{E_n \mid \mathscr{H}_j\}$ represents a (direct) probability (or a probability density), while for a fixed result, for different hypotheses, $\Pr\{E_n \mid \mathscr{H}_j\}$ represents a likelihood. According to Fisher, statistical inferences in cases such as this find their natural expression in terms of likelihood, so that the relevant quantity, if the result E_n has been obtained, is the likelihood ratio

$$\lambda_n = \Pr\{E_n \mid \mathscr{H}_1\}/\Pr\{E_n \mid \mathscr{H}_2\}. \tag{1}$$

If λ_n is greater than 1, $\mathscr{H}_1$ is more likely (on E_n) than $\mathscr{H}_2$, and the magnitude of λ_n measures how much more likely it is; in the converse situation, a similar interpretation is given to $1/\lambda_n$. According to Jeffreys, on the other hand, statistical inferences are to be expressed in terms of (posterior) probability, and in a case such as this one he would say that the ratio of posterior probabilities, or the posterior odds for $\mathscr{H}_1$ against $\mathscr{H}_2$, would be given by

$$\Pr\{\mathscr{H}_1 \mid E_n\}/\Pr\{\mathscr{H}_2 \mid E_n\} = \lambda_n . \lambda_0,$$

where

$$\lambda_0 = \Pr\{\mathscr{H}_1\}/\Pr\{\mathscr{H}_2\}$$

is the ratio of the prior probabilities of $\mathscr{H}_1$ and $\mathscr{H}_2$. In the absence of previous knowledge about the two hypotheses, λ_0 would be taken to be 1, so that the posterior odds would be mathematically the same as the likelihood ratio—the difference between the two would lie in their interpretation only. Finally, according to Neyman and Pearson, the situation could be interpreted in terms of their 'probabilities of error of the first and second kinds', α and β, which are direct probabilities. If it were desired to lay down a discriminating rule for deciding between $\mathscr{H}_1$ and $\mathscr{H}_2$ on the basis of a definite experimental procedure, the result E_n would be regarded as one of a set of possible results which might have been obtained with this procedure; this totality of possible results would then be divided into two sets, C_j, such

that if the result obtained belonged to C_j the hypothesis $\mathscr{H}_j$ would be 'accepted'. The risks of error would then be

$$\alpha = \Pr\{E_n \epsilon C_2 \mid \mathscr{H}_1\} \quad \text{and} \quad \beta = \Pr\{E_n \epsilon C_1 \mid \mathscr{H}_2\}.$$

Now Wald's sequential test procedure is specified by saying that observations are to be taken in stages, and at each stage the likelihood ratio λ_n is to be calculated, and if

$$\alpha'/(1-\beta') < \lambda_n < (1-\alpha')/\beta', \tag{2}$$

then observations are to continue, while they are to stop as soon as either of these two inequalities is broken. Wald has shown that if this test procedure is used, and if the class C_1 is taken as those results for which the right-hand inequality is broken, while C_2 consists of those results for which the left-hand inequality is broken, and if α' and β' are small, then we have, very nearly,

$$\alpha = \alpha' \quad \text{and} \quad \beta = \beta'. \tag{3}$$

Now if, in fact, we have a result for which the likelihood ratio is $\lambda_n\,(>1)$, we may choose to regard it as a terminal result of a sequential procedure in which the limits were set by

$$\lambda_n = (1-\alpha')/\beta',$$

and then we could, using Wald's result, regard λ_n as expressing the ratio of two direct probabilities, the probability of rightly accepting $\mathscr{H}_1$ and the probability of wrongly accepting $\mathscr{H}_1$. A converse interpretation would hold if λ_n were less than 1. Thus in the case of testing two simple hypotheses against each other the three theories of inference can be 'reconciled' in the sense that all three can be taken as saying that the relevant quantity to calculate is λ_n, and all three agree that the larger this quantity is (assuming it greater than 1), the more confident we can be that we would be right to accept $\mathscr{H}_1$. They differ only in that one interprets λ_n as a 'likelihood ratio', another says that λ_n is a ratio of posterior probabilities, while the third interprets λ_n as a ratio of direct probabilities; and this difference comes close to being a merely verbal one. Our object in this paper is to extend this area of 'agreement' between the three theories of inference to cover some of the cases where the two hypotheses being tested against each other are composite. In particular, when we say that a 'likelihood ratio' λ_n has a 'frequency justification', we mean that if it is used as a basis for a Wald sequential procedure (2), then the approximate equalities (3) hold, for the risks of error.

3. The 'sequential t-test' problem we now consider is that of testing against each other the composite hypotheses:

$\mathscr{H}_j\,(j = 1, 2)$: X is normally distributed with standard deviation σ and mean

$$\delta_j \sigma (0 < \sigma < \infty).$$

If $x_i\,(i = 1, 2, \ldots, n)$ represent the results of n independent observations on X, the likelihood function of the observations depends only on t and s, defined by

$$nx. = \sum_i x_i, \quad (n-1)s^2 = \sum_i (x_i - x.)^2, \quad t = x.\sqrt{(n)}/s. \tag{4}$$

Now if $\mathscr{H}_j$ is true about X, it is also true about $Y = aX$, for any $a > 0$; and if the observations are multiplied by a, s becomes as, while t remains unaltered. Thus the likelihood ratio λ_n depends on t only, and is given by

$$\lambda_n = \phi(t \mid \delta_1)/\phi(t \mid \delta_2), \tag{5}$$

where $\phi(t\,|\,\delta)$ is the probability density function of the non-central t distribution with parameter δ and with $(n-1)$ degrees of freedom:

$$\phi(t\,|\,\delta) = K_n \int_0^\infty u^{n-1} \exp -\tfrac{1}{2}\{(n-1)\,u^2 + (ut - \delta\sqrt{n})^2\}\,du$$
$$= K_n \int_0^\infty f_n(u, t\,|\,\delta)\,du, \quad \text{say,} \tag{6}$$

K_n being a constant, depending on n, such that $\int_{-\infty}^{+\infty} \phi(t\,|\,\delta)\,dt = 1$.* We propose now to give a frequency justification of this likelihood ratio, by showing that if the sequential test procedure (2) is followed, using λ_n defined by (5), then the approximate equalities (3) hold for the risks of error α, β, associated with the composite hypotheses $\mathscr{H}_j$.

We notice first that α and β are certainly functions of α' and β',

$$\alpha = \alpha(\alpha', \beta'), \quad \beta = \beta(\alpha', \beta'),$$

and they do not depend on σ. This is so because the criterion λ_n, used in carrying out the test, is unaltered when the observations x_i are replaced by $y_i = ax_i$, for any positive a, i.e. when X is replaced by $Y = aX$ and when σ is replaced by $a\sigma$; to every test on X there is associated a test on Y, leading to the same conclusion, and vice versa, so that the risks of error for X and for Y must be the same.

It follows that in calculating the risks of error we can, if we like, give a particular value to σ in $\mathscr{H}_j$, say, for example, $\sigma = 1$; or, alternatively, we can give σ a distribution of values. The risks of error will be the same in all cases, since they do not depend on σ. In fact, what we shall do is to give σ a distribution of values, in a certain extended sense. We now consider the hypotheses:

$\mathscr{H}_j(a, b)$: X is normally distributed with standard deviation σ and mean $\delta_j\sigma$, and $a < \sigma < b$, with

$$\Pr\{\sigma_1 < \sigma < \sigma_2\} = k \int_{\sigma_1}^{\sigma_2} d\sigma/\sigma, \quad \text{where } k = \log b/a.$$

These are simple hypotheses, in that we can calculate the probability function for the first n observations $x_i\,(i = 1, 2, \ldots, n)$ as a definite number:

$$\Pr\{E_n\,|\,\mathscr{H}_j(a, b)\} = k \int_a^b \{\prod_i (1/[\sqrt{(2\pi)}\,\sigma])\exp -\tfrac{1}{2}(x_i - \delta_j\sigma)^2/\sigma^2\}\,d\sigma/\sigma \prod_i dx_i.$$

Using the relations (4), making the substitution $u = s/\sigma, du/u = -d\sigma/\sigma$ in the integral and cancelling the common factors, we find for the likelihood ratio

$$\lambda_n(a, b) = \Pr\{E_n\,|\,\mathscr{H}_1(a, b)\}/\Pr\{E_n\,|\,\mathscr{H}_2(a, b)\}$$
$$= \int_{sa}^{sb} f_n(u, t\,|\,\delta_1)\,du \Big/ \int_{sa}^{sb} f_n(u, t\,|\,\delta_2)\,du, \tag{7}$$

an expression which would be equal to λ_n if a were zero and b were infinite, s being finite. Now

$$f_n(u, t\,|\,\delta) \leqslant (u \exp -\tfrac{1}{2}u^2)^{n-1},$$

so that as $a \to 0$ and $b \to \infty$ the integrals converge absolutely and uniformly in t and n for $n \geqslant 2$. Further, by taking S_1 small enough and S_2 large enough, we can make $\Pr\{S_1 < s < S_2\}$

* For a practical method of approximating to this likelihood ratio, and for further practical aspects of the test, we refer to Rushton (1950).

as near to 1 as we please for any given σ. Hence, by taking a small enough and b large enough we can, with probability as near as we please to 1, make $\lambda_n(a,b)$ differ from λ_n by as little as we please, uniformly in n and t, for a fixed value of σ. Now from Wald's result referred to in paragraph 2, if $\alpha''(\sigma,a,b)$ and $\beta''(\sigma,a,b)$ are the risks of error using the ratio $\lambda_n(a,b)$ with the test procedure (2) when the standard deviation is σ, we know that

$$(1/\log b/a)\int_a^b \alpha''(\sigma,a,b)\,d\sigma/\sigma = \alpha', \text{ approximately,}$$

and

$$(1/\log b/a)\int_a^b \beta''(\sigma,a,b)\,d\sigma/\sigma = \beta', \text{ approximately.}$$

But the convergence of $\lambda_n(a, b)$ to λ_n, together with the fact that the limits of the inequalities (2) are continuous functions of α' and β' implies that, as $a \to 0$ and $b \to \infty$,

$$\alpha''(\sigma,a,b) \to \alpha \quad \text{and} \quad \beta''(\sigma,a,b) \to \beta,$$

and so we must have, approximately,

$$\alpha = \alpha' \quad \text{and} \quad \beta = \beta'.$$

This completes the frequency justification of the likelihood ratio (5).

4. We now go on to generalize the argument of the preceding paragraph. Suppose we have two composite hypotheses

$$\mathscr{H}_j\text{: } X \text{ has the probability function } \phi(x \mid \gamma_j, \theta),\ \theta \in \Omega,$$

and suppose that

(*a*) Ω can be regarded as a group of transformations on the space of X.

For example, in the special case of §3, the standard deviation is a positive real number, and the set of positive real numbers forms a group under the operation of multiplication; the space of X, in this case, is the real axis, and the group of positive real numbers under multiplication can be regarded as a group of transformations of the real axis into itself—viz. the 'magnification group', corresponding to changes in the scale of measurement. Next we suppose that

(*b*) If X has the probability function $\phi(x \mid \gamma, \theta)$, and $Y = \theta' X$, with $\theta' \in \Omega$, then Y has the probability function $\phi(y \mid \gamma, \theta'\theta)$.

In the case of §3, if X is normally distributed with parameters δ and σ, then $Y = \sigma' X$ is normally distributed with parameters δ and $\sigma'\sigma$. The next supposition is that

(*c*) For any n independent observations $x_i (i = 1, 2, \ldots, n)$ there exist a pair of statistics

$$u_n = u_n(x_1, \ldots, x_n),$$
$$v_n = v_n(x_1, \ldots, x_n),$$

which are jointly sufficient for the parameters γ, θ, so that

$$\prod_i \phi(x_i \mid \gamma, \theta) = f_n(u_n, v_n \mid \gamma, \theta)\, g_n(x_1, \ldots, x_n), \tag{8}$$

and which are such that, for any $\theta' \in \Omega$,

$$u_n(\theta' x_1, \ldots, \theta' x_n) = u_n(x_1, \ldots, x_n),$$

and

$$v_n(\theta' x_1, \ldots, \theta' x_n) = \theta' v_n(x_1, \ldots, x_n).$$

In §3, $u_n = t$ and $v_n = s$, as defined by (4).

Fourthly, we suppose that

(*d*) Ω is a locally compact topological group, containing a sequence of compact sets $\Omega_1 \subset \Omega_2 \subset \ldots \subset \Omega_k \subset \ldots$ such that $\lim_{k\to\infty} \Omega_k = \Omega$.

In the case of §3, the fact that the product of two positive numbers is a continuous function of either of them, shows that the positive real numbers under multiplication form a topological group, and that this group is locally compact is equivalent to the Heine-Borel theorem. A sequence of sets such as Ω_k is obtained by taking Ω_k as the closed interval $(\epsilon_k, 1/\epsilon_k)$, where ϵ_k is the kth element of a sequence of positive numbers which decrease to 0. For definitions of the terms used in assumption (*d*) we refer to Weil (1940), where a concise and elegant account of the theory to be used subsequently in this paper is also given.

Now by a theorem due to A. Haar, if Ω is a locally compact topological group we can define a Borel family $\mathfrak{H}$ of subsets of Ω, containing all the compact sets of Ω, and a measure $\mu(S)$, defined for all $S \in \mathfrak{H}$, such that for any $\theta \in \Omega$,

$$\mu(S) = \mu(\theta S),$$

where θS is the set of images under θ of the elements of S. The measure so defined is called a (left-) invariant Haar measure, and it is unique up to a numerical constant if we exclude trivial cases (e.g. μ identically zero). By means of μ we can define an integral, for every function $f(\theta)$ continuous for $\theta \in S$, with the property

$$\int_{\theta \in S} f(\theta)\, d\mu = \int_{\theta \in \theta' S} f(\theta'^{-1}\theta)\, d\mu \quad \text{for any } \theta' \in \Omega.$$

In the case of §3, the Haar measure of an interval (a, b) on the positive real axis is

$$\mu(a, b) = \int_a^b \frac{dx}{x}.$$

To verify the invariance of this measure, we noticed that the set $c(a, b)$ is the interval (ca, cb), and

$$\begin{aligned} \mu(ca, cb) &= \int_{ca}^{cb} \frac{dx}{x} \\ &= \int_a^b \frac{dy}{y} \quad \text{putting } y = x/c, dy = dx/c \\ &= \mu(a, b). \end{aligned}$$

It is apparent that this particular invariant measure was used to obtain the 'prior distribution' of σ involved in the specification of the hypotheses $\mathscr{H}_j(a, b)$ in §3.

We now give the generalization of this part of the argument.

Consider the pair of simple hypotheses:

$\mathscr{H}_j(k)$: X has the probability function $\phi(x \mid \gamma_j, \theta), \theta \in \Omega_k$, with $\Pr\{\theta \in S\} = \mu(S)/\mu(\Omega_k)$, for $S \subset \Omega_k$.

Using E_n to denote the n independent observations x_i, we shall have, using (8), and omitting a possible factor $\prod_i dx_i$,

$$\Pr\{E_n \mid \mathscr{H}_j(k)\} = \int_{\theta \in \Omega_k} f_n(u_n, v_n \mid \gamma_j, \theta)\, g_n(x_1, \ldots, x_n)\, d\mu / \mu(\Omega_k),$$

so that the likelihood ratio is

$$\lambda_n(k) = \frac{\int_{\theta \in \Omega_k} f_n(u_n, v_n \mid \gamma_1, \theta)\, d\mu}{\int_{\theta \in \Omega_k} f_n(u_n, v_n \mid \gamma_2, \theta)\, d\mu}.$$

We now have to assume that the function f_n decreases sufficiently rapidly outside Ω_k. Specifically, we assume that:

(*e*) As $k \to \infty$, $\lambda_n(k) \to \lambda_n$, where

$$\lambda_n = \frac{\int_{\theta \in \Omega} f_n(u_n, v_n \mid \gamma_1, \theta)\, d\mu}{\int_{\theta \in \Omega} f_n(u_n, v_n \mid \gamma_2, \theta)\, d\mu},$$

and the convergence is uniform in u_n and in n, at least for n sufficiently large.

To complete the generalization we have to show that this λ_n is the ratio of the probability function of u_n on the hypothesis $\mathscr{H}_1$ to the probability function of u_n on the hypothesis $\mathscr{H}_2$.

To find the probability function of u_n we have to integrate the probability function $\prod_i \phi(x_i \mid \gamma, \theta)$ of the sample point $E_n = (x_1, \ldots, x_n)$ over the region of the sample space where u_n = constant. This will give the probability function apart, possibly, from a function of u_n only which will represent the volume element between two neighbouring regions of the sample space corresponding to u_n and $u_n + du_n$; this function of u_n only will cancel when we take the likelihood ratio, so that it can be ignored. Now let us denote by $U(S)$ the region of sample space where $u_n = U$ and where $v_n \in S$; note that S is a subset of Ω, since from condition (*c*) it follows that the space of v_n is Ω. Now for a fixed value of U, the integral

$$\int_{U(S)} g_n(x_1, \ldots, x_n) \prod_i dx_i$$

as a function of S, defines a measure in Ω, and we can therefore set it equal to

$$\int_{v_n \in S} m(U, v_n)\, d\mu.$$

Now using (8) we have

$$\Pr\{E_n \in U(S)\} = \int_{v_n \in S} f_n(U, v_n \mid \gamma, \theta)\, m(U, v_n)\, d\mu. \tag{9}$$

But we also have, from condition (*b*), if we use $\theta_1 E_n$ to denote the sample point $(\theta_1 x_1, \ldots, \theta_1 x_n)$,

$$\Pr\{\theta_1 E_n \in U(\theta_1 S)\} = \int_{v_n \in \theta_1 S} f_n(U, v_n \mid \gamma, \theta_1\theta)\, m(u, v_n)\, d\mu$$

$$= \int_{v_n \in S} f_n(U, \theta_1 v_n \mid \gamma, \theta_1 \theta)\, m(U, \theta_1 v_n)\, d\mu, \tag{10}$$

by the invariant property of the measure μ. The two probabilities (9) and (10), since they correspond to the same event, must be equal. If we remember that the set S is arbitrary, and compare the integrands in (9) and (10), we see that we must have (at least almost everywhere, with respect to μ)

$$f_n(U, \theta_1 v_n \mid \gamma, \theta_1 \theta)\, m(U, \theta_1 v_n) = f_n(U, v_n \mid \gamma, \theta)\, m(U, v_n),$$

which shows that, for fixed U and γ, $f_n(U, v_n \mid \gamma, \theta) m(U, v_n)$ is a function of $\theta^{-1}v_n$ only. Hence the probability function of u_n which is, apart from the function of u_n already noted as irrelevant,

$$\int_{v_n \in \Omega} f_n(u_n, v_n \mid \gamma, \theta) m(u_n, v_n) d\mu$$

is equal to

$$\int_{\theta \in \Omega} f_n(u_n, v_n \mid \gamma, \theta) m(u_n, v_n) d\mu,$$

and when we take the ratio of the two probability functions the factor $m(u_n, v_n)$ will cancel, so that we obtain λ_n as given in (*e*). This completes the generalization of the argument of §3.

As to the conditions (*a*)–(*e*), it may be remarked that the essential ones are (*a*), (*b*) and (*c*). Situations for which the conditions (*d*) and (*e*) are not satisfied may, in theory, occur, but they would be pathological, and the writer has been unable to think of practical situations where the conditions are not satisfied. It should be noted that we have not assumed that X is a scalar variate, so that the argument would apply, for example, to the problem of testing, in a bivariate normal population, whether the correlation coefficient was ρ_1 or ρ_2, the other parameters being unknown; the likelihood ratio would in this case be the ratio of the probability density function of the sample correlation coefficient r on the hypothesis $\rho = \rho_1$ to the same thing on the hypothesis $\rho = \rho_2$.

5. We now turn to the consideration of the point of view of Jeffreys. First, we may note that the likelihood ratio λ_n of §3 would represent, for Jeffreys, the ratio of posterior probabilities of the corresponding hypotheses, since he takes the prior distribution of σ, when nothing is known about it, to be given by the probability element $d\sigma/\sigma$. So that, in this respect, the argument of §2 extends to §3. But this is not all. Jeffreys has not, so far as the writer is aware, laid down general principles which determine the choice of prior distributions in cases of 'ignorance'; but the argument he uses to justify the choice of $d\sigma/\sigma$ strongly suggests that he would, in all cases like that considered in §4, where the conditions (*a*) and (*b*) are satisfied, choose as his prior distribution the appropriate invariant Haar measure. If we assume that this is so, we can say that the argument of §2 extends in this way to the general case of §4.

After the author had developed the argument of §3, while he was working on the generalization of §4, he had a conversation with Charles Stein, from which it became clear that Stein had already worked out the essential features of the generalization, though his mode of presenting the argument differs from mine. Stein is to publish his work, which is along rather different general lines from mine. In spite of our difference of outlook, our conversation was most helpful to me.

REFERENCES

BARNARD, G. A. (1949). Statistical inference. *J.R. Statist. Soc.*, Series B, **11**, 115.

RUSHTON, S. (1950). On a sequential *t*-test. *Biometrika*, **37**, 326.

WEIL, A. (1940). *L'Integration dans les Groupes Topologiques et ses Applications.* Paris: Hermann et Cie. (Actualites Scientifiques et Industrielles.)

EXPERIMENTAL DESIGNS FOR SERIALLY CORRELATED OBSERVATIONS

By R. M. WILLIAMS

Department of Scientific and Industrial Research, Wellington, N.Z.

1. Introduction

In many fields of experiment there is evidence of positive or, more rarely, negative correlation between the errors in adjacent observations; e.g. in yields from plots in a field trial or in the observations on a sequence of articles produced in some industrial process. The most common method (due to R. A. Fisher) of coping with this correlation is to neutralize it by using randomized experimental designs, e.g. randomized blocks or Latin squares. These designs permit of a very simple method of analysis of the observations (analysis of variance). But there are other possible methods, less convenient but perhaps more sensitive, which are worth investigating.

In experiments where any effect such as fertility trend is likely to change fairly smoothly from one plot to the next, and the plots are arranged in a one-dimensional sequence, we can use the systematic designs and the methods of analysis developed by Hald (1948), in which it is assumed either that the fertility trend can be represented by a polynomial or that it is nearly linear, so that it is adequately estimated at any point by a moving average centred on it, the residual errors themselves being uncorrelated.

In this paper we investigate systematic experimental designs which may be used when the observations form a one-dimensional sequence, and we may assume that the errors are correlated in a stationary linear autoregressive process (without trend). Errors of this type may occur in various types of experiment, but for convenience we shall use the terminology of field experiments throughout this paper.

The p treatments are denoted by $\mathbf{1}, \mathbf{2}, \mathbf{3}, \ldots, \mathbf{p}$, and the corresponding treatment effects by $a_1, a_2, \ldots, a_p$. The observations, arranged linearly, are denoted by

$$y_0, y_1, \ldots, y_n,$$

and we assume that they may be represented by

$$y_i = x_i + a_{[i]}, \tag{1}$$

where x_i is the error component and $a_{[i]}$ is the effect of the treatment applied to the ith plot. The notation $[i]$ will be used generally to denote the number of the treatment applied to the ith plot.

In the next section we develop the form of design and analysis when the errors are in the form of a normal stationary Markoff process, i.e.

$$x_i = \rho x_{i-1} + \epsilon_i, \tag{2}$$

where $|\rho| < 1$ and ϵ_i is normally distributed with zero mean and variance σ^2, independently of x_{i-1}, x_{i-2}, etc.

In §3 these results are extended to the second-order case, where

$$x_i + d_1 x_{i-1} + d_2 x_{i-2} = \epsilon_i. \tag{3}$$

2. The design and analysis for the first-order case

2·1. For variables satisfying equation (2) we have var $(x_i) = \sigma^2/(1-\rho^2)$, and the correlation coefficient of any pair of variables (x_i, x_{i+r}) is ρ^r. The joint distribution of $x_0, x_1, \ldots, x_n$ is

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}(n+1)} (1-\rho^2)^{\frac{1}{2}} \exp\left[\frac{-1}{2\sigma^2}\left\{x_0^2+x_n^2+(1+\rho^2)\sum_1^{n-1}(x_i^2)-2\rho\sum_0^{n-1}(x_i x_{i+1})\right\}\right] dx_0 dx_1 \ldots dx_n. \tag{4}$$

For any given design we may obtain the maximum-likelihood equations for $a_1, a_2, \ldots, a_p, \sigma^2, \rho$ from equation (4) in the form

$$\frac{\hat\rho}{(1-\hat\rho^2)} = (n+1)\left[\sum_0^{n-1}(x_i x_{i+1}) - \hat\rho\sum_1^{n-1}(x_i^2)\right]\left[x_0^2+x_n^2+(1+\hat\rho^2)\sum_1^{n-1}(x_i^2)-2\hat\rho\sum_0^{n-1}(x_i x_{i+1})\right]^{-1}, \tag{5}$$

$$\hat\sigma^2 = \frac{(1-\hat\rho^2)}{\hat\rho}\left[\sum_0^{n-1}(x_i x_{i+1}) - \hat\rho\sum_1^{n-1}(x_i^2)\right], \tag{6}$$

$$-\hat\sigma^2\frac{\partial L}{\partial a_r} = (1+\hat\rho^2)\sum_{[i]=r}(y_i-\hat a_r) - \hat\rho\sum_{[i\pm1]=r}(y_i-\hat a_{[i]}) = 0, \tag{7}$$

$$-\hat\sigma^2\frac{\partial L}{\partial a_{[0]}} = (1+\hat\rho^2)\sum_{[i]=[0]}(y_i-\hat a_{[0]}) - \hat\rho\sum_{[i\pm1]=[0]}(y_i-\hat a_{[i]}) - \hat\rho^2(y_0-\hat a_{[0]}) = 0, \tag{8a}$$

$$-\hat\sigma^2\frac{\partial L}{\partial a_{[0]}} = (1+\hat\rho^2)\sum_{[i]=[0]}(y_i-\hat a_{[0]}) - \hat\rho\sum_{[i\pm1]=[0]}(y_i-\hat a_{[i]}) - \hat\rho^2(y_0+y_n-2\hat a_{[0]}) = 0. \tag{8b}$$

Here L is the logarithm of the likelihood. The sum $\sum_{[i]=r}(y_i)$ denotes the sum over all plots receiving **r**, with the convention that if for any i, $[i+1] = [i-1] = r$, the corresponding y_i is included in the sum twice. We use $\hat a_r$, etc., to denote the maximum-likelihood estimators, but for compactness we use x_i, etc., instead of $(y_i - \hat a_{[i]})$, where, as in (5) and (6), no confusion is likely to result. Equation (7) holds for all r other than $[0]$ and $[n]$; if $[0] \neq [n]$ then (8*a*) gives the form for $r = [0]$ and a similar form holds for $r = [n]$; if $[0] = [n]$, both these are replaced by (8*b*).

We may choose the design in any form which simplifies the analysis. In order to reduce the effect of long trends we shall consider only designs composed of m blocks each containing each of the p treatments once, together with, perhaps, one or more extra plots at either end. It will be convenient if the variance of the estimate of treatment effect is the same for all treatments. Two designs that suggest themselves are

I (*a*) $\quad$ **(1, 2, 3, ..., p) (1, 2, 3, ..., p) (1, 2, 3, ..., p) (1, 2, ...)...,**

and

I (*b*) $\quad$ **(1, 2, 3, ..., p) (p, p − 1, ..., 3, 2, 1) (1, 2, 3, ..., p) (p, p − 1, ...)....**

Here each block is enclosed in a bracket for ease of reading. The brackets have no bearing whatever on the conduct of the experiment, which is arranged as a single sequence of mp plots with no break between blocks.

Design I (*a*) is susceptible to linear trends, and in this respect I (*b*) is better, but both are unsatisfactory because the variance of the estimates of treatment differences depends

* Equations (4), (5) and (6) have been given by Koopmans (1942).

strongly on the particular treatments. If y_{ur} is the yield from the plot in the uth block receiving the rth treatment, the solutions of equations (7) and (8) for I (a) are given by

$$\hat{a}_r = y_{.r} - c_1\hat{\rho}^r - c_2\hat{\rho}^{-r},$$

$$c_1 = \frac{1}{m}\left[\frac{\hat{\rho}}{m}(y_{11}-y_{.1}) + \left(\hat{\rho}^p\frac{m-1}{m}-1\right)(y_{mp}-y_{.p})\right]\left[\left(\hat{\rho}^p\frac{m-1}{m}-1\right)^2 - \left(\frac{\hat{\rho}}{m}\right)^2\right]^{-1},$$

$$c_2 = \frac{1}{m}\left[\left(\hat{\rho}^p\frac{m-1}{m}-1\right)(y_{11}-y_{.1}) + \frac{\hat{\rho}}{m}(y_{mp}-y_{.p})\right]\left[\left(\hat{\rho}^p\frac{m-1}{m}-1\right)^2 - \left(\frac{\hat{\rho}}{m}\right)^2\right]^{-1}.$$

If m is large $\hat{a}_r$ tends to $y_{.r}$, and the variance of the estimate of treatment differences is given by

$$\text{var}\,(y_{.r}-y_{.s}) = \frac{2\sigma^2}{m(1-\rho^2)}\left[1-\left(\rho^{|r-s|}+\frac{\rho^p}{1-\rho^p}\left(1-\frac{1}{m}\frac{1-\rho^{mp}}{1-\rho^p}\right)\right)(\rho^{r-s}+\rho^{s-r}-2)\right].$$

For design I (b) we have

$$\hat{a}_r = y_{.r} + \frac{(y_{11}+y_{m1}-2y_{.1})(\hat{\rho}^r+\hat{\rho}^{2p+1-r})}{m\left(1+2\hat{\rho}\dfrac{1}{m}-\hat{\rho}^{2p}\dfrac{m-2}{m}\right)}.$$

Again, if m is large, $\hat{a}_r$ tends to $y_{.r}$ and

$$\text{var}\,(y_{.r}-y_{.s}) = \frac{2\sigma^2}{m(1-\rho^2)}\left[1-\rho^{|r-s|}-\rho^{2p-r-s+1}+\tfrac{1}{2}(\rho^{2p-2r+1}+\rho^{2p-2s+1})\right.$$
$$\left.+\frac{\rho^{2p}}{1-\rho^{2p}}\left(1-\frac{2}{m}\frac{1-\rho^{mp}}{1-\rho^p}\right)\left(1-\frac{\rho^{r-s}+\rho^{s-r}}{2}\right)\left(1-\frac{\rho^{2p-r-s+1}+\rho^{-2p+r+s-1}}{2}\right)\right].$$

A more useful design is found by exploiting the fact that the partial correlation between non-adjacent pairs of plots is zero. To do this we require

(A) that each treatment shall occur equally often (say c times) adjacent to every other treatment, or

(B) that each treatment shall occur equally often adjacent to every treatment including itself.

Such designs will be called II (a) and II (b) respectively. For example,

2(1234)(2314)(3142) and **1(1234)(4132)(2413)(3421)**

satisfy conditions A and B respectively. The construction of these designs is discussed in §2·3.

In order to ensure that condition A or B is fulfilled, an additional plot receiving the treatment applied to the last plot must be placed at the beginning of the set of blocks. (This could be avoided if the arrangement of the plots was circular, but this is not likely to be possible in practice.) This has the further advantage that since the treatments applied to plots o and n are the same, the asymmetry due to end-effects is confined to one treatment and equation (8b) is appropriate. The condition A (or B) ensures that the family of equations given by (7) and (8b) is symmetrical in all the a_r's except that the coefficient of $a_{|0|}$ in (8b) is too large by an amount $(1-\rho^2)$. The remaining asymmetry in (8b) may be eliminated by assuming that the variance of the observation on the first plot is $\sigma^2 h$, where h is large, instead of $(\sigma^2/1-\rho^2)$, but that the distribution of $(x_i-\rho x_{i-1})$ for $i = 1, 2, \ldots$ is otherwise unchanged.* The effect of this is to reduce the weight given to the first plot by

* A similar device is used by Quenouille (1949); the writer is grateful for the opportunity to see that paper before its publication.

a factor ρ^2. If the correlation is high this is negligible, and if very small it is equivalent to ignoring this observation altogether. Equation (8*b*) is now replaced by

$$\frac{\partial L}{\partial a_{[0]}} = (1+\hat{\rho}^2)\left[\frac{y_0\hat{\rho}^2+y_n}{\hat{\rho}^2+1}+\sum_{[i]=[0]}(y_i)-y_0-y_n-m\hat{a}_{[0]}\right]-2\rho\sum_{[i\pm1]=[0]}(y_i-\hat{a}_i)=0, \quad (9)$$

and, provided we adopt the convention that the weighted sum in square brackets is to replace $\sum_{[i]=[0]}(y_i-a_{[0]})$ (which would have $m+1$ terms), we can write this in the same form as equation (7), and we have, for condition A,

$$(1+\hat{\rho}^2)\left[\sum_{[i]=r}(y_i)-m\hat{a}_r\right]-\hat{\rho}\left[\sum_{[i\pm1]=r}(y_i)-c\left(\sum_1^p(\hat{a}_j)-\hat{a}_r\right)\right]=0, \quad (10)$$

and for condition B,

$$(1+\hat{\rho}^2)\left[\sum_{[i]=r}(y_i)-m\hat{a}_r\right]-\hat{\rho}\left[\sum_{[i\pm1]=r}(y_i)-c\sum_1^p(\hat{a}_j)\right]=0. \quad (11)$$

The solution of equation (10) is

$$\hat{a}_r=\frac{(1+\hat{\rho}^2)\sum_{[i]=r}(y_i)-\hat{\rho}\sum_{[i\pm1]=r}(y_i)+\dfrac{2\hat{\rho}\gamma}{p-1}}{m\left(1+\hat{\rho}^2+\dfrac{2\hat{\rho}}{p-1}\right)}, \quad (12)$$

where, by the relations in §2·3,

$$\gamma=\sum_1^n(y_i)+\frac{\hat{\rho}}{1-\hat{\rho}}(y_n-y_0)=m\sum_1^p(\hat{a}_j).$$

Since γ is independent of r it disappears when we estimate treatment differences.

If $\rho\to0$, $\quad \hat{a}_r\to\dfrac{1}{m}\sum_{[i]=r}(y_i)$,

and if $\rho\to1$, $\quad \hat{a}_r\to\dfrac{1}{m}\dfrac{p-1}{p}\left[\sum_{[i]=r}(y_i)-\dfrac{1}{2}\sum_{[i\pm1]=r}(y_i)\right]+\dfrac{\gamma}{mp}$.

The value of $\hat{\rho}$ can be obtained by successive approximations from equations (5) and (12), and $\hat{\sigma}^2$ from equation (6). If ρ is known the maximum-likelihood estimates $\hat{a}_r$ are normally distributed and their variance matrix is given exactly by

$$V^{-1}=\left[-E\left(\frac{\partial^2\log f}{\partial a_r\partial a_s}\right)\right].$$

In practice ρ and σ^2 are both subject to sampling errors, but for large n their distributions are approximately normal, and the variances of these estimates are therefore given to order $1/n$ by inverting the matrix

$$\frac{1}{\sigma^2}\begin{bmatrix} m(1+\rho^2) & -\dfrac{2m}{p-1}\rho & \dots & -\dfrac{2m}{p-1}\rho & 0 & 0 \\ -\dfrac{2m}{p-1}\rho & m(1+\rho^2) & \dots & -\dfrac{2m}{p-1}\rho & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -\dfrac{2m}{p-1}\rho & -\dfrac{2m}{p-1}\rho & \dots & m(1+\rho^2) & 0 & 0 \\ 0 & 0 & \dots & 0 & \dfrac{n\sigma^2}{1-\rho^2} & 0 \\ 0 & 0 & \dots & 0 & 0 & \dfrac{n}{2\sigma^2} \end{bmatrix},$$

which leads to

$$\left.\begin{aligned}
\operatorname{var}(\hat{a}_r) &= \frac{\sigma^2}{m}\frac{(p-1)(1-\rho)^2+2\rho}{((1+\rho^2)(p-1)+2\rho)(1-\rho)^2},\\
\operatorname{cov}(\hat{a}_r\hat{a}_s) &= \frac{\sigma^2}{m}\frac{2\rho}{((1+\rho^2)(p-1)+2\rho)(1-\rho)^2},\\
\operatorname{var}(\hat{a}_r-\hat{a}_s) &= \frac{2\sigma^2}{m}\frac{1}{\left(1+\rho^2+\dfrac{2\rho}{p-1}\right)} = V_{\mathrm{II}\,(a)},\\
\operatorname{var}(\hat{\rho}) &= \frac{1-\rho^2}{n},\\
\operatorname{var}(\hat{\sigma}^2) &= \frac{2\sigma^4}{n},\\
\operatorname{cov}(\hat{\rho},\hat{\sigma}^2) &= \operatorname{cov}(\hat{\rho}\hat{a}_r) = \operatorname{cov}(\hat{\sigma}^2\hat{a}_r) = 0.
\end{aligned}\right\} \tag{13}$$

For design II (b) we have from equation (11)

$$\hat{a}_r = \frac{(1+\hat{\rho}^2)\sum\limits_{[i]=r}(y_i)-\hat{\rho}\sum\limits_{[i\pm1]=r}(y_i)+\dfrac{2\hat{\rho}\gamma}{p}}{m(1+\hat{\rho}^2)}. \tag{14}$$

Here, if $\rho\to 0$,

$$\hat{a}_r \to \frac{1}{m}\sum_{[i]=r}(y_i),$$

and if $\rho\to 1$,

$$\hat{a}_r \to \frac{1}{m}\left[\sum_{[i]=r}(y_i)-\frac{1}{2}\sum_{[i\pm1]=r}(y_i)\right]+\frac{\gamma}{mp}.$$

Also

$$\left.\begin{aligned}
\operatorname{var}(\hat{a}_r) &= \frac{\sigma^2}{m}\frac{m(1+\rho^2)-(p-1)\rho c}{(1+\rho^2)(m(1+\rho^2)-p\rho c)},\\
\operatorname{cov}(\hat{a}_r\hat{a}_s) &= \frac{\sigma^2}{m}\frac{\rho c}{(1+\rho^2)(m(1+\rho^2)-p\rho c)},\\
\operatorname{var}(\hat{a}_r-\hat{a}_s) &= \frac{2\sigma^2}{m(1+\rho^2)} = V_{\mathrm{II}\,(b)}.
\end{aligned}\right\} \tag{15}$$

$\operatorname{Cov}(\hat{a}_r\hat{\rho})$, $\operatorname{cov}(\hat{a}_r\hat{\sigma}^2)$, $\operatorname{cov}(\hat{\rho}\hat{\sigma}^2)$, $\operatorname{var}(\hat{\rho})$ and $\operatorname{var}(\hat{\sigma}^2)$ are the same as for type II (a).

2·2. *The efficiency of designs II (a) and II (b).* If, instead of a design of type II (a) or II (b) a randomized block design is applied to a set of observations arranged linearly in m blocks of p treatments, in which the variance of each observation is σ_1^2 and the correlation between x_i and $x_{i+\nu}$ is ρ_ν, then the treatment difference is estimated by the difference of the means of the plots receiving those treatments, and the variance of this difference is given by

$$2\sigma_1^2(1-\bar{\rho})/m,$$

where $\bar{\rho}$ is the average correlation between any two plots in the same block, taken over the $p!$ possible arrangements.

For a general stationary process

$$\bar{\rho} = \frac{2}{p(p-1)}\sum_{\nu=1}^{p-1}(p-\nu)\rho_\nu,$$

and in the case of the linear Markoff process where $\rho_\nu = \rho^\nu$ and $\sigma_1^2 = \sigma^2/(1-\rho^2)$, the variance is

$$\frac{2\sigma^2}{m(1-\rho^2)}\left[1-\frac{2\rho}{p-1}\left(\frac{1}{1-\rho}-\frac{1}{p}\frac{(1-\rho^p)}{(1-\rho)^2}\right)\right] = V_R.$$

The efficiency of the systematic designs compared with the randomized scheme is given by $V_R/V_{\mathrm{II}(a)}$ and $V_R/V_{\mathrm{II}(b)}$. These are given in Tables 1 and 2 for $p = 2, 3, 4, 10, \infty$.

Table 1. $V_R/V_{\mathrm{II}(a)}$

ρ \ p	2	3	4	5	10	∞
1	2·00	2·000	2·222	2·500	4·070	∞
0·9	1·90	1·854	2·010	2·208	3·202	9·526
0·8	1·80	1·717	1·817	1·953	2·564	4·556
0·7	1·70	1·589	1·648	1·733	2·096	2·922
0·6	1·60	1·470	1·496	1·545	1·749	2·125
0·5	1·50	1·361	1·363	1·387	1·492	1·667
0·4	1·40	1·263	1·250	1·257	1·303	1·381
0·3	1·30	1·183	1·161	1·158	1·169	1·198
0·2	1·20	1·108	1·088	1·082	1·080	1·083
0·1	1·10	1·043	1·029	1·025	1·020	1·020
0	1·00	1·000	1·000	1·000	1·000	1·000
−0·1	0·90	0·977	0·977	1·010	1·016	1·020
−0·2	0·80	0·980	1·028	1·034	1·072	1·083
−0·3	0·70	1·016	1·100	1·134	1·178	1·198
−0·4	0·60	1·098	1·231	1·283	1·350	1·381
−0·5	0·50	1·250	1·451	1·525	1·624	1·667
−0·6	0·40	1·520	1·824	1·925	2·067	2·125
−0·7	0·30	2·019	2·496	2·632	2·842	2·922
−0·8	0·20	3·080	3·910	4·097	4·443	4·556
−0·9	0·10	6·370	8·288	8·569	9·335	9·520
−1·0	0	∞	∞	∞	∞	∞

Table 2. $V_R/V_{\mathrm{II}(b)}$

ρ \ p	2	3	4	5	10	∞
1	1	1·333	1·667	2·000	3·667	∞
0·9	0·953	1·238	1·510	1·768	2·884	9·526
0·8	0·911	1·154	1·371	1·570	2·313	4·556
0·7	0·876	1·081	1·255	1·404	1·898	2·922
0·6	0·850	1·020	1·156	1·266	1·593	2·125
0·5	0·833	0·972	1·077	1·156	1·370	1·667
0·4	0·829	0·939	1·017	1·072	1·211	1·381
0·3	0·838	0·927	0·981	1·018	1·102	1·198
0·2	0·867	0·929	0·964	0·987	1·036	1·083
0·1	0·918	0·949	0·966	0·976	0·998	1·020
0	1·000	1·000	1·000	1·000	1·000	1·000
−0·1	1·122	1·085	1·068	1·062	1·039	1·020
−0·2	1·300	1·213	1·179	1·144	1·120	1·083
−0·3	1·557	1·401	1·347	1·315	1·255	1·198
−0·4	1·933	1·676	1·598	1·550	1·462	1·381
−0·5	2·500	2·083	1·979	1·906	1·782	1·667
−0·6	3·400	2·720	2·584	2·470	2·291	2·125
−0·7	4·967	3·808	3·634	3·440	3·174	2·922
−0·8	8·200	6·013	5·795	5·418	4·983	4·556
−0·9	18·100	12·670	12·398	11·405	10·494	9·520
−1·0	∞	∞	∞	∞	∞	∞

For large positive or negative correlations there is a considerable increase in efficiency for all values of $p > 2$. The curious behaviour of design II (a) for $p = 2$ is understandable when it is realized that the only such design is of the form

$$2(12)(12)(12)....$$

Such a design is very inefficient for negative correlations.

The relative efficiency of II (a) and II (b) is given by

$$\frac{V_{\text{II}(b)}}{V_{\text{II}(a)}} = 1 + \frac{2\rho}{(p-1)(1+\rho^2)}.$$

For positive correlations II (a) and for negative correlations II (b) is more efficient. The difference is important if p is small and $|\rho|$ approaches unity.

2·3. *The construction of type II designs.* In order to satisfy the conditions of design II (a) II (b), m the number of blocks, p the number of treatments and c the number of times a given treatment is adjacent to another, must satisfy the relations

$$2m = c(p-1) \quad \text{for II } (a)$$

or

$$2m = cp \quad \text{for II } (b).$$

For II (b) we have the further restriction that c must be even, since if a treatment occurs in adjacent plots the first is adjacent to the second and the second adjacent to the first, so that c can only change by multiples of two.

If it is necessary to keep the design as short as possible we can put $c = 1$, and since for this, or any odd value of c, the condition on II (a) requires that p shall be odd, we have to introduce an extra treatment if p is even. If c is even we may use II (a) if the correlation between adjacent plots is expected to be positive, II (b) if it is negative.

A design may be made up by taking a design for say, $c = 1$, and repeating this design in a continuous chain; or we may repeat the design but rename the treatments, randomly or systematically, at each repetition, subject to the restrictions that the treatments applied to the first and last plots should be the same in each repetition. Alternatively, we can make up a design for, say, $p = 5$, $c = 7$, composed of one design for $c = 1$, followed by three different designs for $c = 2$, subject to the same restriction as above; or we can make a design in which the condition A (§ 2·1) is satisfied for the design as a whole, but not for any part of it. This last is laborious to construct and is not likely to be of general use, and most designs are likely to be made up of a sequence of basic designs for which c is 1 or 2.

If each repetition of the basic design is independent of the others (e.g. in an experiment on animals, if each repetition is applied to a different animal), then the permutations can be made without the restriction on the first and last treatments.

No general algebraic method has been found for the construction of these designs, but a number of them have been constructed fairly easily by means of the diagram in Fig. 1. The treatments are arranged round a circle and the design traced out by joining the treatments by a continuous line so that all possible joins are made c times.

For designs of type II (b) we can represent the passage from one plot to an adjacent one receiving the same treatment by one or more closed loops at each vertex according as c is 2, 4, etc.

If we relax the condition that the blocks should contain each treatment once and only once then we can apply to this figure the theorem, due to Euler, that a figure which has no

odd node can be described unicursally in a re-entrant route (see, for example, Rouse Ball's *Mathematical Essays and Recreations*, p. 173). This shows that if c is even designs of both types exist for p odd or even; and such designs can be constructed by means of Trémaux's rule for threading a maze (ibid. p. 183). If c is odd the only designs that exist are type II (a) for p odd; Trémaux's rule does not then apply.

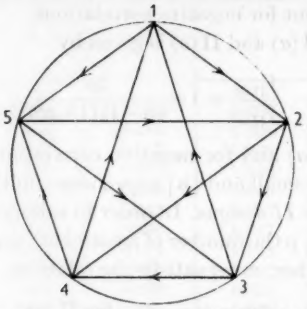

Fig. 1. Construction of type II(a) design (12345)(24153) with $p=5$, $c=1$.

If the above condition is imposed, this rule may fail because the only join still to be made at some point is to a treatment which has already appeared in the block. It has, however, been found in practice that this situation is easily avoided.

Examples of designs constructed in this manner for $c = 1, 2, p = 3, 4, 5, \ldots, 10$, are given below.

II (a)

$c = 1$

(123)
(12345) (24135)
(1234567) (2461357) (3625147)
(123456789) (246813579) (369471582) (591483726)

$c = 2$

(12)
(123) (213)
(1234) (2314) (3142)
(12345) (34152) (45312) (31425)
(123456) (253641) (531642) (516324) (314562)
(1234567) (5364721) (4251637) (1354762) (5643271) (6241573)
(12345678) (64251387) (53627481) (73416582) (63748125) (32458671) (64138275)
(123456789) (759864231) (473852691) (827163594) (876925143) (947563821) (354681927) (397162485)
(123456789 10) (685724913 10) (28364795 10 1) (48 10 7392615) (243517698 10) (418596 10 327) (36475 10 9281) (938456217 10) (497861 10 253)

II (b)

$c = 2$

(12) (21)
(123) (312) (231)
(1234) (4132) (2413) (3421)
(12345) (53142) (25134) (45123) (35241)
(123456) (613524) (462315) (524163) (346512) (263541)
(1234567) (7246135) (5147362) (2571364) (4576123) (3471526) (6537241)
(12345678) (81357246) (68251473) (36178452) (26713485) (56812374) (46153827) (75836241)
(123456789) (913572468) (814792563) (371582694) (483957126) (613458927) (781469325) (591683742) (284976351)
(123456789 10) (10 246813579) (9147 10 36258) (8271593 10 64) (4 10 16923875) (5 10 84937162) (21 10 9854376) (65 10 8427913) (35149286 10 7) (74 10 2596381)

3. THE EXTENSION TO THE SECOND ORDER PROCESS

The relation between the errors instead of being $x_i = \rho x_{i-1} + \epsilon_i$ may be of the form

$$x_i + d_1 x_{i-1} + d_2 x_{i-2} = \epsilon_i, \tag{16}$$

where again ϵ_i is normally distributed about zero with variance σ^2. We have the following well-known results:

(i) The process is stationary if, and only if, $-1 < d_2 < 1$ and $-(1+d_2) < d_1 < (1+d_2)$.

(ii) The variance of x_i is

$$\frac{\sigma^2(1+\xi_1\xi_2)}{(1-\xi_1\xi_2)(1-\xi_1^2)(1-\xi_2^2)} = \frac{\sigma^2(1+d_2)}{(1-d_2)((1+d_2)^2-d_1^2)},$$

where ξ_1, ξ_2 are the roots of $\xi^2 + \xi d_1 + d_2 = 0$.

(iii) The correlation ρ_ν between x_i and $x_{i+\nu}$ is given by

$$\rho_\nu + d_1\rho_{\nu-1} + d_2\rho_{\nu-2} = 0, \tag{17}$$

so that

$$\rho_\nu = c_1\xi_1^\nu + c_2\xi_2^\nu,$$

where

$$c_1 = \frac{\xi_1(1-\xi_2^2)}{(\xi_1-\xi_2)(1+\xi_1\xi_2)},$$

$$c_2 = \frac{\xi_2(1-\xi_1^2)}{(\xi_2-\xi_1)(1+\xi_1\xi_2)}.$$

Two problems now arise:

(i) Can a design be found which will simplify the analysis as in the first order case?

(ii) What errors will be introduced if the design is analysed on the assumption that it is a first order process, when in fact it is the second process?

Some designs do exist which simplify the analysis. These are obtained by imposing simultaneously two conditions on the design:

(1) That each treatment should occur equally often adjacent to every other treatment.

(2) That each treatment should occur equally often adjacent but one to every other treatment.

Such designs will be called type III and must fulfil the condition for II(*a*) that $2m = c(p-1)$ (designs corresponding to II(*b*) do not exist). The construction of such designs involves a good deal of trial and error, but is made easier by the simultaneous use of two of the circular diagrams described above.

Examples of such designs are:

23(123)(123),
24(1234)(2134)(1324),
43(12345)(24153)(14532)(51243).

In this case it is necessary to repeat the last two plots at the beginning of the design in order to satisfy conditions (1) and (2). If it is assumed that the variance of these two observations (say $h\sigma^2$) is large compared with σ^2, we can eliminate the asymmetry due to end effects.

If the errors in the observations are named $x_{-1}, x_0, x_1, x_2, \ldots, x_n$, then the distribution will be approximately

$$\frac{1}{h}\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}(n+2)} \exp\left[-\frac{1}{2\sigma^2}\sum_{1}^{n}(x_i + d_1x_{i-1} + d_2x_{i-2})^2\right]\prod_{-1}^{n}(dx_i). \tag{18}$$

The maximum-likelihood equations are:

$$\left.\begin{aligned}
\frac{\partial L}{\partial d_1} &= -\frac{1}{\sigma^2}\left[d_1\sum_1^n (x_{i-1}^2) + d_2\sum_1^n (x_{i-1}x_{i-2}) + \sum_1^n (x_{i-1}x_i)\right] = 0,\\
\frac{\partial L}{\partial d_2} &= -\frac{1}{\sigma^2}\left[d_1\sum_1^n (x_{i-1}x_{i-2}) + d_2\sum_1^n (x_{i-2}^2) + \sum_1^n (x_{i-2}x_i)\right] = 0,\\
\frac{\partial L}{\partial \sigma^2} &= -\left(\frac{n+2}{2}\right)\frac{1}{\sigma^2} - \frac{1}{2\sigma^4}\sum_1^n (x_i + d_1x_{i-1} + d_2x_{i-2})^2 = 0.
\end{aligned}\right\} \tag{19}$$

The general expression for $\partial L/\partial a_r (r \neq [n-1], [n])$ is

$$\frac{\partial L}{\partial a_r} = -\frac{1}{\sigma^2}[(1+d_1^2+d_2^2)\sum_{[i]=r}(x_i) + (d_1+d_1d_2)\sum_{[i\pm1]=r}(x_i) + d_2\sum_{[i\pm2]=r}(x_i)] = 0, \tag{20}$$

and, as in the first order case, $\partial L/\partial a_{[n-1]}$ and $\partial L/\partial a_{[n]}$ can be written in this form provided we take appropriately weighted means for some of the terms involved. These weights can easily be obtained from the equations.

With this interpretation, the maximum-likelihood estimates of a_r are

$$\hat{a}_r = \frac{1}{m\left(1+\hat{d}_1^2+\hat{d}_2^2-\frac{2}{p-1}(\hat{d}_1+\hat{d}_2+\hat{d}_1\hat{d}_2)\right)}\left[(1+\hat{d}_1^2+\hat{d}_2^2)\sum_{[i]=r}(y_i) + (\hat{d}_1+\hat{d}_1\hat{d}_2)\sum_{[i\pm1]=r}(y_i) + \hat{d}_2\sum_{[i\pm2]=r}(y_i) - \frac{2m}{p-1}(\hat{d}_1+\hat{d}_2+\hat{d}_1\hat{d}_2)\sum_1^p(\hat{a}_j)\right], \tag{21}$$

where $$m\sum_1^p(\hat{a}_j) = \sum_1^n(y_i) + \frac{(y_{n-1}-y_{n-1})(\hat{d}_1\hat{d}_2+\hat{d}_2^2) + (y_0-y_n)(\hat{d}_1^2+\hat{d}_1\hat{d}_2+\hat{d}_2^2)}{(1+\hat{d}_1+\hat{d}_2)^2}.$$

The solution of equation (19) can be obtained by successive approximation as before. The term in $\sum_1^p(\hat{a}_j)$ is required only in the estimation of d_1, d_2, and not directly in comparing treatment differences.

The variances, obtained as in § 2, are (to order $1/n$)

$$\left.\begin{aligned}
\operatorname{var}(\hat{a}_r) &= \frac{\sigma^2}{m}\frac{1+d_1^2+d_2^2+2\frac{p-2}{p-1}(d_1+d_2+d_1d_2)}{\left(1+d_1^2+d_2^2-\frac{2}{p-1}(d_1+d_2+d_1d_2)\right)(1+d_1+d_2)^2},\\
\operatorname{cov}(\hat{a}_r\hat{a}_s) &= \frac{\sigma^2}{m}\frac{\frac{-2}{p-1}(d_1+d_2+d_1d_2)}{\left(1+d_1^2+d_2^2-\frac{2}{p-1}(d_1+d_2+d_1d_2)\right)(1+d_1+d_2)^2},\\
\operatorname{var}(\hat{a}_r-\hat{a}_s) &= \frac{2\sigma^2}{m}\frac{1}{1+d_1^2+d_2^2-\frac{2}{p-1}(d_1+d_2+d_1d_2)},\\
\operatorname{var}(\hat{d}_1) &= \operatorname{var}(\hat{d}_2) = (1-d_2^2)/n,\\
\operatorname{var}(\hat{\sigma}^2) &= 2\sigma^4/n,\\
\operatorname{cov}(\hat{d}_1\hat{d}_2) &= d_1(1-d_2)/n,\\
\operatorname{cov}(\hat{\sigma}^2\hat{d}_1) &= \operatorname{cov}(\hat{\sigma}^2\hat{d}_2) = \operatorname{cov}(\hat{\sigma}^2\hat{a}_r) = \operatorname{cov}(\hat{d}_1\hat{a}_r) = \operatorname{cov}(\hat{d}_2\hat{a}_r) = 0.
\end{aligned}\right\} \tag{22}$$

The efficiency of these designs compared with a randomized scheme is shown in Table 3 for $p = 5$, for various values of d_1, d_2.

Table 3

d_1 \ d_2	−0·9	−0·6	−0·3	0	0·3	0·6	0·9
1·8	—	—	—	—	—	—	0·35 62·80 181·50
1·5	—	—	—	—	—	0·47 15·68 33·08	0·46 17·35 37·88
1·2	—	—	—	—	0·62 8·96 14·32	0·64 4·37 6·82	0·61 8·84 14·40
0·9	—	—	—	0·74 6·30 8·57	0·86 2·52 2·94	0·87 2·75 3·16	0·76 8·11 10·67
0·6	—	—	0·72 4·88 6·79	0·94 1·82 1·92	1·10 1·64 1·50	1·06 2·14 2·01	0·87 6·86 7·89
0·3	—	0·59 4·04 6·81	0·82 1·45 1·78	1·06 1·21 1·13	1·20 1·31 1·09	1·10 1·88 1·71	0·86 6·36 7·41
0·0	0·44 3·49 7·88	0·60 1·22 2·03	0·81 0·99 1·23	1·00 1·00 1·00	1·06 1·19 1·11	0·94 1·79 1·90	0·74 6·28 8·54
−0·3	—	0·55 0·86 1·56	0·70 0·84 1·21	0·81 0·93 1·16	0·82 1·16 1·42	0·72 1·82 2·53	0·58 6·54 11·35
−0·6	—	—	0·55 0·80 1·45	0·60 0·94 1·54	0·59 1·20 2·03	0·53 1·92 3·65	0·44 7·06 16·16
−0·9	—	—	—	0·44 0·98 2·21	0·43 1·29 3·02	0·39 2·10 5·44	0·33 7·84 23·69
−1·2	—	—	—	—	0·32 1·42 4·48	0·29 2·44 8·45	0·26 8·94 33·74
−1·5	—	—	—	—	—	0·22 2·64 11·89	0·20 10·14 50·96
−1·8	—	—	—	—	—	—	0·16 11·50 72·33

The numbers given are:

(i) $(m/2\sigma^2) \times$ variance of the maximum-likelihood estimator using systematic designs of type III.

(ii) $(m/2\sigma^2) \times$ variance using randomized designs.

(iii) The efficiency of type III designs compared with the randomized scheme, i.e. (ii)/(i).

In the points where no entries are made the process is not stationary. When the process is nearly unstable, the correlation is very high, and the variance of the randomized block estimates is consequently high. This accounts for the high value of the efficiency in these cases.

4. The effect of deviations from the autoregressive form

4·1. The assumption that the error is due to a linear stochastic process is at best an approximation. We shall only consider the effect of deviations from this assumption in the first-order case. Two possible sources of error are:

(i) That a monotonic trend is present.

(ii) That the correlation does not obey the law $\rho_v = \rho^v$.

4·2. The effect of a monotonic trend on the estimate of treatment differences is reduced by the condition that the treatments are arranged in blocks, and designs can be constructed so that the mean position of each treatment is approximately the same. Where the design consists of a number of repetitions of the same basic design we can achieve this by permuting the treatments subject to the restrictions of §2·3. In the case of designs of type II (*b*) this can be done by reversing every second repetition.

A trend is therefore not likely to give a serious bias to the estimate of the treatment differences, but a monotonic trend will lead to an overestimate of the value of the correlation ρ. This will tend to reduce the estimate of the variance of treatment differences given in equations (13) and (15) except for design II (*a*) when ρ is less than $-1/(p-1)$, in which case the estimate will be increased. On the other hand, since the weights in equations (12) and (14) are incorrect, the actual variances of the treatment differences will be increased. With the above exception, the effect of these errors will be cumulative in leading us to underestimate the variance.

4·3. To estimate the importance of deviations from the rule $\rho_v = \rho^v$, we consider the effect of applying the design and analysis appropriate to the first-order process, when the errors are in fact due to a second-order process.

If the actual relation is $x_i + d_1 x_{i-1} + d_2 x_{i-2} = \epsilon_i$ and the relation is assumed to be $x_i = \rho x_{i-1} + \epsilon_i$, then, since the maximum-likelihood estimator of ρ is approximately equal to the first serial correlation coefficient, ρ will be estimated as $\rho_1 = -d_1/(1+d_2)$.

The treatment difference $(a_r - a_s)$ will be estimated from equation (12) using $\rho = \rho_1$, and we denote it by $(a_r - a_s)'$. This estimate is unbiased.

The actual variance V_1' of $(a_r - a_s)'$ will always be greater than the variance V_2 of the estimate based on the second-order hypothesis. We also have an inaccurate estimate V_1 of the variance of $(a_r - a_s)'$ derived from equation (13).

The variance V_1' of $(a_r - a_s)'$ will depend not only on the parameters d_1, d_2, σ^2, but also on the pairs of values r, s and on the particular design chosen. The calculations are laborious, and to simplify them we shall calculate only the average value of V_1' over all pairs r, s and we shall assume

(i) That the design consists of l repetitions of a basic design containing q blocks of p treatments, so that $m = lq$.

(ii) That end-effects can be ignored.

(iii) That the parameters are not subject to sampling errors.

The assumption (i) makes it possible to reduce the detailed calculation of V_1' to the calculation of

$$\operatorname{var}\Big[\sum_{[i]=r,B}(x_i)-\sum_{[i]=s,B}(x_i)\Big]=\operatorname{var}(x_i)\sum_{t=0}^{pq}\beta_t\rho_t,$$

where $\sum_{[i]=r,B}(x_i)$ is the sum over the basic design and $\sum_{[i]=r,L}(x_i)$ is the sum over the l repetitions of the basic design. If

$$\phi_{1l}=l\xi_1^l+\xi_1^{pq}\left(\frac{l(1-\xi_1^{pq})-(1-\xi_1^{lpq})}{(1-\xi_1^{pq})^2}\right)(\xi_1^l+\xi_1^{-l}),$$

and ϕ_{2l} is the same function in ξ_2, then we have

$$\left.\begin{aligned}
&\operatorname{var}\Big[\sum_{[i]=r,L}(x_i)-\sum_{[i]=s,L}(x_i)\Big]=\operatorname{var}(x_i)\sum_0^{pq}\beta_t(c_1\phi_{1l}+c_2\phi_{2l}),\\
&\operatorname{cov}\Big[\Big(\sum_{[i]=r,L}(x_i)-\sum_{[i]=s,L}(x_i)\Big)\Big(\sum_{[i\pm1]=r,L}(x_i)-\sum_{[i\pm1]=s,L}(x_i)\Big)\Big]\\
&\qquad=\operatorname{var}(x_i)\sum_0^{pq}\beta_t[c_1(\xi_1+\xi_1^{-1})\phi_{1l}+c_2(\xi_2+\xi_2^{-1})\phi_{2l}],\\
&\operatorname{var}\Big[\sum_{[i\pm1]=r,L}(x_i)-\sum_{[i\pm1]=s,L}(x_i)\Big]=\operatorname{var}(x_i)\left[\frac{l\beta_0(1-\xi_1^2)(1-\xi_2^2)(1-\xi_1\xi_2)}{\xi_1\xi_2(1+\xi_1\xi_2)}\right.\\
&\qquad\left.+\sum_0^{pq}\beta_t(c_1(\xi_1+\xi_1^{-1})^2\phi_{1l}+c_2(\xi_1+\xi_2^{-1})^2\phi_{2l})\right].
\end{aligned}\right\}\tag{23}$$

Applying these results to equation (12) and noting that $\beta_0=2q$, we have

$$V_1'=\frac{\sigma^2}{m^2\left[d_1^2+(1+d_2)^2-\dfrac{2d_1(1+d_2)}{(p-1)}\right]^2}\left[\frac{2m(1+d_2)^2d_1^2}{d_2}-\frac{d_2^2\sum_0^{pq}\beta_t((\xi_1-\xi_1^{-1})^3\phi_{1l}-(\xi_2-\xi_2^{-1})^3\phi_{2l})}{(1-d_2)(\xi_1-\xi_2)}\right].\tag{24}$$

Here only β_t depends on the particular design and the values of (r,s). The average of the variance (written $\overline{\operatorname{var}}(a_r-a_s)'=\bar{V}_1'$) over all pairs (r,s) is given by replacing β_t by their average values $\bar{\beta}_t$ given by

$$\operatorname{var}(x_i)\sum_{t=0}^{pq}(\bar{\beta}_t\rho_t)\equiv\frac{2}{p(p-1)}\left[p\sum_{j=1}^{p}\operatorname{var}\sum_{[i]=j,B}(x_i)-\operatorname{var}(x_i)\left(pq+2\sum_{t=0}^{pq}(pq-t)\rho_t\right)\right].$$

If $l=1$, the coefficients of β_t in equations (23) and (24) may be more easily expressed directly in terms of $\rho_t,\rho_{t\pm1},\rho_{t\pm2}$, which can be obtained directly from equation (17). If $l\neq1$, equation (24) must be used.

If $|\xi_1|$, $|\xi_2|$ are small, the variance is inversely proportional to l, and the results in Table 4 show that even for quite large values of ξ_1, ξ_2, this approximation is fairly good.

The value of V_1 would be estimated from equation (13) putting

$$\rho=\rho_1=-d_1/(1+d_2),$$

and

$$\sigma^2=(1-\rho_1^2)\operatorname{var}(x_i)$$

$$=(1-\rho_1^2)\frac{\sigma^2(1+d_2)}{((1+d_2)^2-d_1^2)(1-d_2)}.$$

These results were applied to the series

(*a*) $$x_i - 0{\cdot}9x_{i-1} + 0{\cdot}5x_{i-2} = \epsilon_i,$$

(*b*) $$x_i - 0{\cdot}3x_{i-1} - 0{\cdot}5x_{i-2} = \epsilon_i,$$

(*c*) $$x_i - 0{\cdot}6x_{i-1} = \epsilon_i.$$

For all series $$\rho_1 = 0{\cdot}6, \quad \text{var}\,(\epsilon_i) = \sigma^2.$$

A type III design

(12345) (24153) (14532) (51243)

was chosen so that a correct maximum-likelihood estimator would be available for comparison. The variance of this estimator V_2, and the variance of the estimate under randomization, i.e. $V_R = (2/m)\,\text{var}\,(x_i)\,(1-\bar{\rho})$ (see § 2·2) are also given for comparison.

Table 4

	(*a*)	(*b*)	(*c*)
$\frac{\bar{V}_1'}{\sigma^2}$	$\frac{0{\cdot}266l - 0{\cdot}011(l-1)}{l^2}$	$\frac{0{\cdot}498l + 0{\cdot}003(l-1)}{l^2}$	$\frac{0{\cdot}301}{l}$
$\frac{V_1}{\sigma^2}$	$\frac{0{\cdot}402}{l}$	$\frac{0{\cdot}402}{l}$	$\frac{0{\cdot}301}{l}$
$\frac{V_2}{\sigma^2}$	$\frac{0{\cdot}201}{l}$	$\frac{0{\cdot}300}{l}$	$\frac{0{\cdot}301}{l}$
$\frac{V_R}{\sigma^2}$	$\frac{0{\cdot}849}{l}$	$\frac{0{\cdot}423}{l}$	$\frac{0{\cdot}466}{l}$

These results suggest that in series such as (*b*), where the correlogram differs appreciably from that of a first-order process, fairly serious errors can be introduced and the error underestimated; and it is important that series deviations from the first-order model should be detected.

5. Application to experimental data

Two sets of uniformity trial data were tested to see whether the variability could be described by a linear autoregressive model.

To test the goodness of fit we used the test given by Quenouille (1947), of which the principal result is that if the autoregressive relation is

$$x_i + d_1 x_{i-1} + \ldots + d_k x_{i-k} = \epsilon_i,$$

and if $$(1 + d_1 u + d_2 u^2 + \ldots + d_k u_k)^2 = \sum_{-\infty}^{\infty} A_i u^i,$$

$$R_s = \sum_{-\infty}^{\infty} A_i r_{s-i} \quad (s = k+1, k+2, \ldots),$$

where r_j is the sample serial correlation coefficient, then, if the sample number n is large, the forms R_s are approximately normally distributed with zero mean and variance $\sum_{-\infty}^{\infty} \frac{A_i X_i}{n-s}$ where $X_i = \sum_{-\infty}^{\infty} \rho_\nu \rho_{\nu-i}$. The forms $R_s (s > k)$ are distributed independently of each other and of

$r_s - \rho_s (s \leqslant k)$. The values of $d_1, d_2, \ldots, d_k$ may be determined from the first k values of r_s and the goodness of fit tested by using the forms R_s.

We denote the form R_s when a first- or second-order process has been fitted by $R_{s(1)}$ or $R_{s(2)}$ respectively.

We tested data kindly supplied by Rothamsted Experimental Station giving the yields obtained each year from 1852 to 1946 from two of the plots (referred to as plot 2B which received farmyard manure, and plot 8 which received artificial manure each year) of the Broadbalk wheat experiment, in which wheat has been grown on the same plots each year, receiving the same manurial treatment throughout the experiment. Taking the deviations from the mean yield for the whole period the first serial correlation coefficient r_1 was found to be significant at the 1 % point for plot 2B and at the 5 % level for plot 8.

The values of r_s, R_s for both series are given in Table 5.

Table 5

s	Plot 2B		Plot 8		
	r_s	$R_{s(1)}$	r_s	$R_{s(1)}$	$R_{s(2)}$
1	0·3788	—	0·2162	—	—
2	0·2112	0·0677	0·2265	0·1797	—
3	0·1519	0·0463	0·1008	0·0130	0·0203
4	0·2140	0·1292	0·2387	0·2057	0·1748
5	0·2862	0·1459	0·2556	0·1571	0·1596
6	0·2501	0·0640	0·3526	0·2532	0·1950
7	0·1366	−0·0118	0·0461	−0·0944	−0·1468
8	0·0746	0·0070	0·1088	0·1053	−0·0041
9	0·1076	0·0707	0·0870	0·0421	0·0653
10	0·1939	0·1231	0·2129	0·1804	0·1603
11	0·1067	−0·0248	0·0832	−0·0964	−0·0128
12	0·0535	0·0005	0·0151	−0·0109	−0·0782
13	−0·0018	−0·0270	0·0044	0·0018	−0·0125
14	−0·0651	−0·0560	0·1173	0·1161	0·1236
15	−0·1527	−0·1034	−0·1201	−0·1682	−0·1588
16	−0·1496	−0·0433	−0·0982	−0·0408	−0·0959
17	−0·1444	−0·0530	−0·0973	−0·0604	−0·0133
18	−0·2379	−0·1500	−0·1681	−0·1306	−0·1037
19	−0·1612	−0·0017	−0·0556	0·0625	0·0763
20	−0·0151	0·0729	−0·0349	−0·0403	0·0154
var (R_s)		0·00780	—	0·00967	0·00909
$\chi^2 = \sum_{2}^{20} \frac{R_s^2}{\text{var}\,(R_s)}$		14·97	—	30·72	24·11

For plot 2B the value of χ^2 is not significant. There is a tendency for the positive and negative value to occur together, but the number of runs of positive and negative values is not significant at the 5 % level (see Swed & Eisenhart, 1943), and the form $x_i - 0{\cdot}3788x_{i-1} = \epsilon_i$ seems to fit reasonably well. For plot 8 the value of χ^2 for the first-order process is just significant at the 5 % level; a second-order process $x_i - 0{\cdot}1754x_{i-1} - 0{\cdot}1886x_{i-2} = \epsilon_i$ gives a satisfactory fit. The residuals obtained after fitting these series were tested for randomness and found to be not significant.

The data used to test whether the autoregressive model could be fitted to a series of agricultural plots arranged in a line were taken from the uniformity trial data given by Garber, McIlvaine & Hoover (1926, p. 263, column 1). Each plot, which was sown with wheat, measured 68×21 ft. There were only forty-five consecutive plots available. This series is really too short to be satisfactory, but in the absence of other suitable data we made use of it, while recognizing that not too much weight should be given to the result.

The observations, expressed in bushels per acre as deviations from the mean are given in Table 6.

Table 6

Plot	Yield	Plot	Yield	Plot	Yield	Plot	Yield
1	−0·5	13	0·9	25	−2·0	37	−2·4
2	−6·0	14	1·0	26	−0·8	38	−5·7
3	−0·6	15	4·0	27	1·9	39	−0·9
4	−3·1	16	2·8	28	1·2	40	−0·2
5	−1·3	17	0·4	29	4·1	41	0·5
6	1·7	18	0·6	30	1·7	42	−0·8
7	−0·1	19	−4·8	31	11·1	43	−0·7
8	−2·2	20	−4·7	32	10·9	44	−4·0
9	3·5	21	−5·2	33	10·9	45	−0·4
10	3·5	22	−2·1	34	6·0		
11	−4·9	23	2·6	35	5·2		
12	0·7	24	−3·5	36	1·6		

Table 7

s	r_s	$R_{s(1)}(n-s)^{\frac{1}{2}}$
1	0·626	—
2	0·430	0·249
3	0·267	−0·169
4	0·182	0·102
5	−0·109	−1·467
6	−0·138	0·437
7	−0·286	−0·962
8	−0·224	0·487
9	−0·262	−0·564
10	−0·384	−0·852
11	−0·508	−0·758
12	−0·509	−0·138

A first-order form, $x_j - 0{\cdot}626x_{i-1} = \epsilon_i$, was fitted, and the goodness of fit tested by comparing $\sum_{2}^{12} (R^2_{s(1)}(n-s))$ (see Table 7) with its estimated variance, $(1-r_1^2)^2$. The value of 14·18 so obtained is not significant for 11 degrees of freedom.

We conclude that, as far as Quenouille's test can be trusted for such a small value of n, the data appear to be adequately represented by a first-order process.

SUMMARY

A method of experimental design is developed for use when the errors are in the form of a first- or second-order linear autoregressive stochastic process. Estimates of treatment effects and the efficiency relative to a randomized block design applied to the same data are given. The effect of deviations from the first-order process on the efficiency of the design is dealt with. Consideration is given to the suitability of these designs for agricultural data.

I wish to acknowledge gratefully my indebtedness to Mr F. J. Anscombe for suggesting this subject of research and for his advice and assistance during my investigations at the Statistical Laboratory, Cambridge. I also wish to thank the Department of Scientific and Industrial Research, New Zealand, for financial assistance at that time.

REFERENCES

GARBER, R. J., MCILVAINE, T. C. & HOOVER, M. M. (1926). A study in soil heterogeneity in experimental plots. *J. Agric. Res.* **33**, 255.

HALD, A. (1948). The decomposition of a series of observations. *G.E.C. Gads Forlag.* Copenhagen.

KOOPMANS, T. (1942). Serial correlation and quadratic forms in normal variables. *Ann. Math. Statist.* **13**, 14.

QUENOUILLE, M. H. (1947). A large-sample test for the goodness of fit of autoregressive schemes. *J.R. Statist. Soc.* **110**, 123.

QUENOUILLE, M. H. (1949). On a method of trend elimination. *Biometrika*, **36**, 75.

SWED, F. S. & EISENHART, C. (1943). Tables for testing randomness of grouping in a sequence of alternatives. *Ann. Math. Statist.* **14**, 66.

THE TIME INTERVALS BETWEEN INDUSTRIAL ACCIDENTS

By B. A. MAGUIRE, E. S. PEARSON and A. H. A. WYNN

1. Introduction

Statistical methods have been used for many years for studying the frequency of occurrence of accidents in fixed intervals of time. Statisticians are familiar with the use by Bortkiewicz (1898) of the Poisson distribution to analyse the numbers of men in ten Prussian army corps killed by the kick of a horse, quoted, for example, by Yule (1922) and still being quoted in text-books (Jeffreys, 1948). Again, the series of papers describing research on industrial accidents published by the Industrial Health Research Board, which began with a paper by Greenwood & Woods (1919), are based on the analysis of the frequency of accidents occurring in fixed intervals of time.

It is, however, a characteristic of an accident that it occurs at a particular instant of time, which is often recorded. Thus the basic data consist of a sequence of ordered intervals of varying length. Analysis may therefore be applied to the time intervals between accidents rather than to the frequencies of accidents occurring in successive fixed intervals of time. Which of the alternative methods of analysis is the most efficient will depend on the type of question we are asking and the assumptions regarding the data which we are prepared to make. Under certain conditions the two methods of attack are clearly equivalent. Thus if we are prepared to assume that accidents are taking place at random in time and at a constant average rate, we shall obtain the same estimate of this rate* either from the average number of accidents occurring in successive fixed time intervals or from the average length of the varying interval between the accidents.

When, however, a more detailed analysis is required, as, for example, in testing for changes with time, it seems likely (provided accurate interval data are available) that methods of analysis based on these will be more powerful than those often employed using only the accident frequencies in relatively long fixed intervals of time. This advantage will lie partly in the fact that for an analysis to be sensitive to changes in time it must be able to handle observations in small groups, and an exact distribution theory for the continuous variable (the interval) is available, where it does not exist in manageable form for the discontinuous variable (the frequency or count). In the present paper, however, we are not concerned with a comparison of the efficiency of the two methods of attack, but rather with an examination of available methods of handling the interval analysis.

Any conclusion about the accident liability of a single individual must always be based on very small numbers even in unusually dangerous occupations. Serious accidents of particular kinds, such as explosions in mines, are fortunately infrequent. The extent to which any body of accident statistics may be usefully analysed depends in part upon the power of the statistical methods to generalize from these small numbers. The loss of information entailed in compiling accident frequencies in fixed intervals of time may, therefore, seriously reduce the usefulness of the records.

* Apart from a marginal difference due to the fact that the time stretch considered may not be precisely the same in the two cases.

In routine accident control it is important for a manager to have the earliest possible indication of a significant change in the expectation of accident. Administrative measures may have been taken to reduce accidents; it is important to have the earliest possible indication of the effect of these measures. Analysis of the time intervals between accidents may provide an earlier indication of an improvement than any analysis of accident frequency.

Failure in the past to use time-interval analysis for research on industrial accidents has been partly due to the more complicated records necessary and to the difficulty of defining intervals. This paper is not concerned with these practical difficulties but introduces a discussion of the many statistical techniques which are available for the analysis of time-interval data if the practical difficulties can be overcome.

There is one note of caution which should be raised. If we take for consideration the sample of intervals between accidents occurring in a given period of time, e.g. in a year, the data are in some degree selected. For instance, no intervals of a year or more can be included in the sample. While this should not affect the theory seriously, provided the average interval is short compared to the fixed period, the matter is one requiring investigation.

2. THE EXPONENTIAL DISTRIBUTION

It was shown by Whitworth (1901) that if the expectation of events per unit time is constant, then the time intervals between events are exponentially distributed. If E is the expectation of accidents per unit time, then the probability density function of time intervals may be written

$$f(t) = E e^{-Et}. \tag{1}$$

For this distribution, both mean and standard deviation equal $1/E$.

Table 1. *Time intervals in days between explosions in mines, involving more than 10 men killed, from 6 December 1875 to 29 May 1951*

378	286	871	66
36	114	48	291
15	108	123	4
31	188	457	369
215	233	498	338
11	28	49	336
137	22	131	19
4	61	182	329
15	78	255	330
72	99	195	312
96	326	224	171
124	275	566	145
50	54	390	75
120	217	72	364
203	113	228	37
176	32	271	19
55	23	208	156
93	151	517	47
59	361	1613	129
315	312	54	1630
59	354	326	29
61	58	1312	217
1	275	348	7
13	78	745	18
189	17	217	1357
345	1205	120	(complete interval
20	644	275	to 29 May 1951)
81	467	20	

Mean time interval = 241 days.

Table 1 gives the time intervals between explosions in mines in Great Britain involving the loss of ten lives or more since 1875. In Fig. 1 these intervals are shown to be approximately exponentially distributed. A study of other sequences of industrial accidents shows that the time intervals are usually exponentially distributed to at least a rough approximation.

Table 2. *Time intervals in days between successive compensable accidents in one district of a mine*

3	2	4	8	3
23	0	0	2	12
0	0	0	0	10
0	2	2	1	4
2	3	1	1	0
3	0	1	8	3
0	0	1	0	2
2	0	1	0	2
1	0	8	0	0
0	0	0	1	0
1	3	0	8	14
0	4	2	3	1
0	2	0	0	4
0	3	2	2	8
0	5	2	0	
1	0	0	1	
0	0	2	1	
1	2	0	3	
3	4	0	2	
2	5	2	0	
1	0	8	4	

Mean time interval = 2·2653 days.

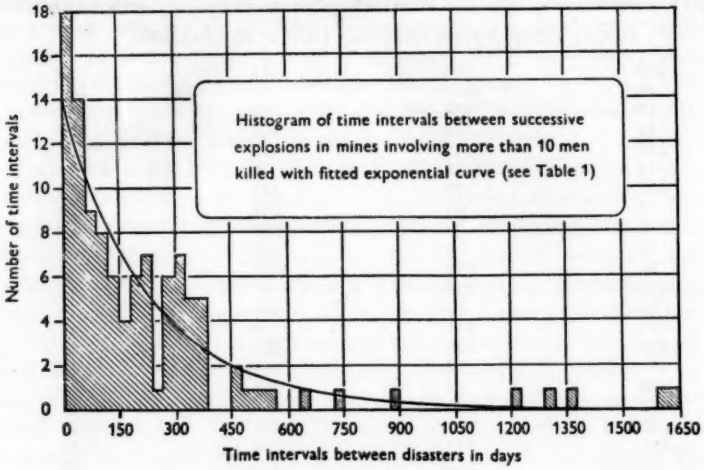

Fig. 1.

Table 2 gives the time interval between successive accidents for one shift in a section of a mine. Fig. 2 shows that the distribution is approximately exponential.

If the risk of accident were constant, the exponential distribution would be a good fit. In order to show that there have been changes in risk or expectation of accident, tests of

homogeneity or goodness of fit are required for an exponential distribution. This method of attack has been suggested by various writers including Bortkiewicz (1898), Morant (1920), Neyman & Pearson (1928) and Sukhatme (1936).

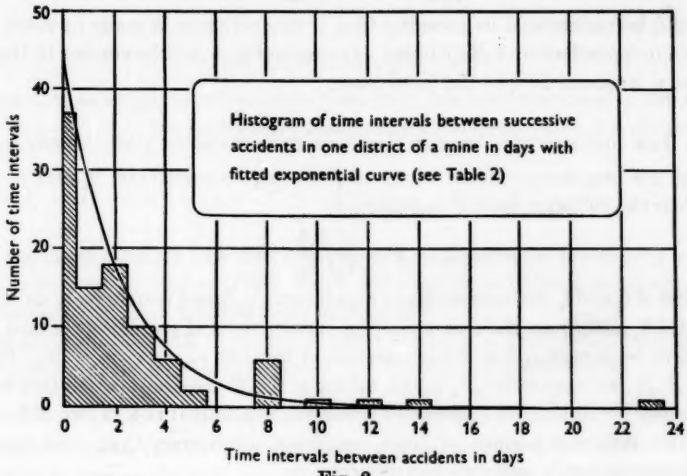

Fig. 2.

The exponential distribution has also been used in a more generalized form which assumes that following each event there is a 'closed period', t_0, during which another event cannot occur; (1) then becomes

$$f(t) = E\,e^{-E(t-t_0)}. \tag{2}$$

This distribution has been applied by various writers, including Sukhatme (1936), to the analysis of the time intervals between telephone calls. No useful application to the study of industrial accidents appears, however, to have been discovered so far, and the simpler distribution (1), where $t_0 = 0$, appears adequate for accident analysis.

3. The distribution of the sum and mean of time intervals

It may be shown that if time intervals are exponentially distributed, then both the sum, T, and the mean, $\bar{t}$, of n intervals have a Pearson type III distribution. The distribution of the sum of n intervals, say T, is

$$f(T) = \frac{(ET)^{n-1}\,e^{-ET}\,E}{(n-1)!}. \tag{3}$$

The distribution of the mean of n intervals, say $\bar{t}$, is

$$f(\bar{t}) = \frac{(nE\bar{t})^{n-1}\,e^{-nE\bar{t}}\,nE}{(n-1)!}. \tag{4}$$

The distribution (4) has a mean of $1/E$ and standard deviation of $1/(\sqrt{(n)}\,E)$; as n becomes large, this distribution tends to the normal form. $\bar{t}$ is also a sufficient estimator of $1/E$.

It follows from the form of equations (1) and (4) that $2Et$ and $2nE\bar{t}$ are distributed as χ^2 with degrees of freedom $\nu = 2$ and $\nu = 2n$, respectively. This means that well-known properties of the χ^2-distribution and of related functions, as well as existing tables, may be

employed in accident interval analysis. For example, $\bar{t}$ may be used to provide confidence limits for an unknown value of E. Thus if $\chi^2_{1-\alpha}$ and χ^2_α are the lower and upper 100α % points of the distribution of χ^2, having $\nu = 2n$ degrees of freedom, there is a probability of $1-2\alpha$ that

$$\chi^2_{1-\alpha}/(2n\bar{t}) \leqslant E \leqslant \chi^2_\alpha/(2n\bar{t}).$$

This should be understood as meaning that if this estimate is made in cases where the intervals are independent and distributed exponentially, it will be correct in the long run, whatever be n, in about $100(1-2\alpha)$ % of cases.

4. The distribution of the ratio of the means of time intervals

If χ^2_1 and χ^2_2 are two independent values of χ^2 having, respectively, ν_1 and ν_2 degrees of freedom, then the variance ratio F is defined as

$$F = \frac{\chi^2_1}{\nu_1} \Big/ \frac{\chi^2_2}{\nu_2}.$$

It follows that if $\bar{t}_1$ and $\bar{t}_2$ are independent mean intervals, based respectively on n_1 and n_2 intervals, and if E_1 and E_2 are the corresponding expectations of accidents per unit time, then $E_1\bar{t}_1/(E_2\bar{t}_2)$ will be distributed as F with degrees of freedom $\nu_1 = 2n_1, \nu_2 = 2n_2$. Thus to test whether $E_1 = E_2$, we may refer $\bar{t}_1/\bar{t}_2$ to the tables of the F-distribution. In this way we can test very simply for significant differences between the accident risk in two different places or during two different periods of time, assuming, of course, that time intervals are distributed independently and exponentially.

Thus when operations started on a particular working face in a mine there were seven compensable accidents with a mean time interval of 32 days. New machinery was then introduced and there were twelve further accidents with a mean time interval of 21 days. The ratio $32/21 = 1{\cdot}52$. The 10 % point for F is 1·86 for $\nu_1 = 2\times(7-1) = 12$ and $\nu_2 = 2\times(12-1) = 22$ degrees of freedom. The evidence from these data that the introduction of machinery has increased the expectation of accident is thus very weak.

As indicated in the preceding section, we may also obtain confidence limits for E or $1/E$ based, say, on the value of $\bar{t} = 21$ days observed after new machinery had been introduced. For $\nu = 22$ the lower and upper 5% points for χ^2 are 12·34 and 33·92 respectively. Hence we may state that

$$\frac{22\times 21}{33{\cdot}92} \leqslant \frac{1}{E} \leqslant \frac{22\times 21}{12{\cdot}34} \quad \text{or} \quad 13{\cdot}6 \leqslant \frac{1}{E} \leqslant 37{\cdot}4 \text{ days}$$

with a probability of $1-2\times 0{\cdot}05 = 0{\cdot}90$ of being correct, provided that the intervals follow the exponential law.

As a further illustration, consider the following case. A district in a mine had been working for over two years and there had been sixty-three compensable accidents. The mean time interval was 13·7 days. It became necessary to transfer a group of men to a new district replacing the experienced men by new face workers. Immediately after the change there were five accidents with a mean interval of 1·8 days. The ratio of the means is $F = 7{\cdot}6$; the numbers of degrees of freedom for F are 124 and 8, which give a 1% point of about 4·9. The change resulted, therefore, in a highly significant increase in the expectation of accident.

5. The extreme observations in samples from an exponential distribution

The χ^2 and F-distribution may both be usefully applied to the analysis of time intervals between accidents. Their application depends, however, on the assumption of homogeneity.

Particular tests for homogeneity may be based on the distribution of the extreme intervals. Range tests are particularly useful for indicating significantly long periods of immunity from accident. Thus in Table 1 it will be noted that there are intervals of 1613 and 1630 days between disasters. Is there evidence that the risk of disaster was lower during these periods?

(*a*) *Distribution of* $t_n/(n\bar{t})$

If t_n is the largest among n independent time intervals and $\bar{t}$ the mean of the n intervals, then if we let $g = t_n/(n\bar{t})$, the probability that g exceeds a given value G is shown by Fisher (1929) to be

$$\Pr\{g > G\} = n(1-G)^{n-1} - \tfrac{1}{2}n(n-1)(1-2G)^{n-1} + \ldots + (-1)^{k-1}\frac{n!}{k!(n-k)!}(1-kG)^{n-1},$$

where k is the largest integer less than $1/G$. 5 and 1 % significance levels for g up to $n = 50$ have been tabulated by Fisher (1929, 1950), who has shown that a good approximation can often be obtained from the first term of the series.

Application to data in Table 1. Here $n = 109$, $\bar{t} = 241$, $t_n = 1630$. The significant value of g at the 5 % significance level is $g_{\cdot 05} = 0{\cdot}068703$. Hence t_n is significant if

$$t_n \geqslant g_{\cdot 05} \times n\bar{t} = 1805 \text{ days}.$$

Our longest interval, 1630, is therefore not significant at the 5 % level.

Application to data in Table 2. $n = 98$, $n\bar{t} = 222$, $t_n = 23$. We find that $g = 0{\cdot}1036$, and from the formula above, of which the first term is sufficient,

$$\Pr\{g > 0{\cdot}1036\} = 0{\cdot}0024,$$

so that having regard to the mean of 2·27 days, the largest interval of 27 days is clearly significant.

(*b*) *Distribution of* t_n/t_1

Hartley (1951) has recently considered the ratio, $F_{\max.} = s^2_{\max.}/s^2_{\min.}$, where $s^2_{\max.}$ and $s^2_{\min.}$ are the largest and smallest out of k independent mean-square estimates of variance each based on the same number of degrees of freedom ν. He has also shown that when $\nu = 2$

$$\Pr\{F_{\max.} \leqslant F\} = k\int_0^\infty e^{-x}(e^{-x} - e^{-Fx})^{k-1}\,dx. \tag{5}$$

Clearly if t_n and t_1 are the largest and smallest of n time intervals, their ratio t_n/t_1 will follow Hartley's distribution (5). It must be remembered, however, that relatively small inaccuracies due to recording the shortest interval to the nearest day, week or month have a large influence on the value of t_n/t_1.

In Table 1 the ratio $t_n/t_1 = 1630/1$, which is not significant at the 5 % level. However, since t_1, recorded to the nearest day, is given in this case as 1, we can only say that $0{\cdot}5 < t_1 < 1{\cdot}5$; thus $F_{\max.}$ may lie between 3260 and 1087. Indeed the shortest time interval may in practical cases often be recorded as zero, as in the list of intervals in Table 2 between successive accidents in one district of a mine. It is therefore not likely that the ratio t_n/t_1 will be found very useful in this field of application.

The distributions of t_n and $t_n - t_1$ may be useful if there is a satisfactory basis for assuming a hypothetical value for the expectation of accidents.

(c) *Distribution of* t_n

We have for t_n

$$\Pr\{t_n > \tau\} = 1-(1-e^{-E\tau})^n, \tag{6}$$

where E is the expectation of accidents per unit time.

Application to data in Table 1. $n = 109, E = 0{\cdot}004150$. We find from (6) that with 109 observations the 5 % significance level of Et_n is 7·5268. Thus if the true value of E were as found for this 75-year period, the 5 % significance level for the largest interval would be 1813 days. The observed value, $t_n = 1630$, falls within this limit. It also falls within the 10 % limit, calculated on the same basis.

Application to data in Table 2. $n = 98$, $E = 0{\cdot}4414$. The 5 % significance level of Et_n is 7·4866. Using the observed E, this gives 16·96 as the significant value of t_n. The largest interval, 23, falls beyond this level.

This test as applied in these illustrations, is of course not exact, as E has been determined from the data.

(d) *Distribution of* $t_n - t_1$

For the variate $y = t_n - t_1$, we have

$$\Pr\{y > Y\} = \int_Y^\infty (n-1)\,Ee^{-EY}(1-e^{-EY})^{n-2}\,dY$$

$$= 1-(1-e^{-EY})^{n-1}, \tag{7}$$

a result very similar to (6) except that $n-1$ replaces n. If n is large these tests are very nearly the same. For n small it may be that the test based on $t_n - t_1$ is more powerful.

6. The M-test

Tests may be applied to discover whether there is any significant tendency for intervals to succeed one another in groups, sometimes longer, sometimes shorter. A failure of the overall distribution of intervals to correspond with the exponential might be due to such a tendency; within short periods the exponential law might hold but E might change from period to period. The whole sequence may be broken into k groups of n successive intervals and may be tested for a significant difference between the k mean intervals, these mean intervals being each estimates of $1/E$.

If $\bar{t}_i$ is the mean interval in the ith group of n consecutive intervals ($i = 1, 2, \ldots, k$), where the intervals are exponentially distributed with parameter E_i, then the probability density function for $\bar{t}_i$ is given by

$$f(\bar{t}_i)\,d\bar{t}_i = \frac{1}{\Gamma(n)}(nE_i\bar{t}_i)^{n-1}e^{-n\bar{t}_iE_i}\,d(nE_i\bar{t}_i) \tag{8}$$

or $2nE_i\bar{t}_i$ is distributed like χ^2 with $\nu = 2n$ degrees of freedom, i.e. the mean intervals $\bar{t}_1, \bar{t}_2, \ldots, \bar{t}_k$ of the k groups of n successive intervals will, on our hypothesis, be distributed independently as

$$\chi^2/(2E_in).$$

If the data are homogeneous, the E_i will be equal. This can be examined by using the standard test for homogeneity of variances in samples from a normal population, commonly called Bartlett's test, though in the case of equal groups it is identical with the test given by Neyman and Pearson in 1931.*

* See note on p. 180 added in proof.

The test criterion to be calculated is*

$$M = 2nk\left\{\log_e\left(\frac{1}{k}\sum_{i=1}^{k}\bar{t}_i\right) - \frac{1}{k}\sum_{i=1}^{k}(\log_e \bar{t}_i)\right\}$$
$$= 2nk\left\{\log_e\left(\frac{1}{k}\sum_i T_i\right) - \frac{1}{k}\sum_i \log_e T_i\right\}, \tag{9}$$

where $T_i = n\bar{t}_i$ is the sum of n intervals in the ith group. Bartlett (1937) shows that M/C is distributed approximately as χ^2 with $\nu = k-1$, where

$$C = 1+(k+1)/(6nk). \tag{10}$$

For n small and $k \leqslant 15$, the tables computed by Thompson & Merrington (1946) give the 5 and 1 % points for M.

If ν is large we may have to make use of the fact that, to a good approximation, $\sqrt{(2\chi^2)} - \sqrt{(2\nu-1)}$ is a normal deviate with unit standard deviation. The M-test may also be used when the numbers of intervals in each group are unequal, the necessary modifications to equation (9) being given in the papers quoted.

The data of Table 2 were broken up into fourteen consecutive groups of $n = 7$ intervals, the fourteen sums, T_i, calculated and M obtained from equation (9) as follows:

$$M = 196\,(\log_{10} 2{\cdot}2653 - \tfrac{1}{14}3{\cdot}99028)\log_e 10$$
$$= 31{\cdot}639.$$

Further, $C = 1+15/(6\times 98) = 1{\cdot}0255$.

Hence $M/C = 30{\cdot}85$ with 13 degrees of freedom. Regarded as χ^2, M/C is just significant at the 0·5 % level. There appears therefore to be evidence of fluctuations in accident expectation with time during the period considered.

The M-test may also be applied to single intervals, i.e. when $n = 1$, but here, in practice, the application may be impossible if the record gives zero intervals making some values of $\log_e T_i = -\infty$. Table 1 contains no zeros, and we find that the set of single intervals is just significant at the 10 % level, M being 155·589. Thus the evidence that E is varying is slight.

It will be noticed that the results of applying the M-test in this section confirm those obtained with the g-test in §5 above; the intervals in Table 1 are consistent with a constant accident expectation while those in Table 2 are not.

If a sequence of intervals is broken up into small groups in the way suggested and M calculated, we are perhaps using the best method available of testing whether the intervals follow a common exponential law and are arranged in random order. The test does, however, involve an arbitrary element in the sense that the division points separating the group of n intervals and the value of n itself are at the statistician's choice. In the ideal test this element would be eliminated; but to derive such a test more thought may be needed as to the kind of alternatives to homogeneity which may be expected to occur.

7. A FURTHER EXAMPLE

We shall take as an example to which several methods of analysis may be applied some data for the intervals between accidents causing fatalities in the mines of Great Britain. The seven lists in Table 3 correspond to the seven divisions of the National Coal Board and

* In Bartlett's (1937) notation, $M = -2\log_e \mu$.

the units of time are days. The period covered is 245 days in 1950. The first and last intervals in each division are not intervals between accidents, but between the start of the period and the first accident and between the last accident and the end of the period; they must therefore be omitted from any interval analysis.

Table 3

Div. 1	Div. 2	Div. 3	Div. 4	Div. 5	Div. 6	Div. 7
16 2	4 13 8	12 0	16	21 4	7 3	17 2 0
3 2	9 4 4	1 13	8	2 5	9 20	2 1 1
1 6	0 0 1	7 2	24	15 1	10 7	4 7 1
0 14	11 3 3	1 13	6	1 13	4 12	1 4 15
7 10	1 4 2	4 7	5	5 6	2 0	2 1 2
7 2	12 4 3	5 5	14	1 9	6 23	0 0 5
20 2	0 15 4	9 11	2	9 3	9 13	6 5 3
6 2	0 0 3	6 0	11	1 —	14 19	1 3 14
1 7	1 5 0	5 0	8	0 —	0 12	4 6 2
20 3	2 8 5	8 3	16	17 —	7 —	2 6 5
5 15	26 3 0	2 8	8	0 —	0 —	0 0 24
10 13	2 3 2	13 5	98	1 —	5 —	0 1 0
4 4	4 1 13	8 1	1	24 —	1 —	14 7 1
6 8	4 0 2	1 13	3	14 —	14 —	2 3 0
19 —	6 1 3	1 18	16	4 —	2 —	10 5 8
1 —	0 0 0	4 17	9	9 —	9 —	2 2 —
9 —	5 5 3	3 5	—	20 —	8 —	5 0 —
16 —	0 3 0	6 —	—	14 —	2 —	9 0 —
0 —	0 1 9	11 —	—	1 —	5 —	6 0 —
2 —	2 4 0	11 —	—	1 —	10 —	13 2 —
2 —	5 2 7	6 —	—	44 —	12 —	8 1 —

If we are prepared to assume that accidents occur randomly in time and that the accident expectation was constant within a division during the 245 days, and merely want to test whether there is evidence that the rate differs between divisions, we may apply the simple χ^2 test to the total number of accidents. Here we assume that the frequency of accidents in a fixed period of time is a Poisson variable. The frequency of accidents in each division, $n_j, j = 1, 2, \ldots, 7$, is given in Table 4, and we find that

$$\chi^2 = \sum_{j=1}^{7} (n_j - \bar{n})^2 \Big/ \bar{n} = 44{\cdot}23 \text{ with } \nu = 6,$$

where $\bar{n} = 37{\cdot}14$ is the average number of accidents. This result is clearly very significant so that we must conclude that the differences among the n_j are real.*

Allowing, therefore, E to vary between divisions, we may ask whether the intervals within a division appear to follow the exponential law of equation (1), with values of E_j which may differ. A rapid check on this is obtained from Fisher's g-test, the necessary information being given in Table 4, where as before t_n is the largest of n intervals in a division, T is the sum of the intervals and $g = t_n/T$.† The number of intervals for the jth division is, of course, $n_j - 1$.

For $n_j - 1 < 50$, 5 and 1 % significance levels are shown, taken from Fisher's (1950) table. For the other two cases, the probability level corresponding to the observed g is shown. The

* Even when Division 4 with its exceptional interval of 98 days is omitted, $\chi^2 = 26{\cdot}22$ with $\nu = 5$ is still a highly significant value.

† The suffices j, which strictly should be given, have been dropped for simplicity.

only exceptional result is for Division 4, where g exceeds the 1 % level; this draws attention to the single long interval of 98 days. None of the other six values of g are significant at the 5 % level. The exponential distribution has, of course, a very long 'tail', and the test is mainly helping us to determine whether apparently outlying observations are, in fact, exceptional. To derive a more critical test of departure from the exponential, we must specify the form of departure to which it is wished that the test should be sensitive.

The M-test may also be applied to these data. Allowing for between-divisional differences in E_j, it searches for fluctuations in accident risk during the course of the 245-day period. Omitting the first and last intervals, we have summed the intervals within each division into k consecutive groups of $n = 4$, omitting the last intervals where the total number was not a multiple of 4. The results are shown in Table 5, where M and the Bartlett corrective factor C have been defined in equations (9) and (10). The only value of χ^2 which can be regarded as significant is that for Division 7 which falls just beyond the 0·5 % point. When the 7 values of χ^2 are summed, we find that for $\nu = 54$

$$\Pr\{\chi^2 > 74{\cdot}20\} = 0{\cdot}036.$$

Table 4

Division (j)	n_j	t_n	T	g	Significance levels 5 %	Significance levels 1 %	
1	34	20	221	0·0905	0·1835	0·2237	
2	62	26	234	0·1111	—	—	$\Pr\{g > 0{\cdot}1111\} = 0{\cdot}057$
3	37	18	228	0·0789	0·1712	0·2086	
4	15	98	220	0·4455	0·3517	0·4272	
5	27	44	221	0·1991	0·2212	0·2699	
6	29	23	226	0·1018	0·2088	0·2547	
7	56	24	220	0·1091	—	—	$\Pr\{g > 0{\cdot}1091\} = 0{\cdot}105$

Table 5

Division	k	M	C	$\chi^2 = M/C$	$\nu = k - 1$
1	8	5·41	1·047	5·17	7
2	15	18·66	1·044	17·87	14
3	9	8·12	1·046	7·76	8
4	3	3·88	1·056	3·67	2
5	6	6·92	1·049	6·60	5
6	7	4·50	1·048	4·30	6
7	13	30·13	1·045	28·83	12
			Total	74·20	54

Without Division 7, the χ^2's sum to 45·32 which for 42 degrees of freedom is not exceptional.

Have we any justification in concluding that the intervals for Division 7 show significant changes in risk? We are faced here with the arbitrary character referred to above of our selection of the consecutive group of four intervals. If instead of taking the groups

2, 4, 1, 2), (0, 6, 1, 4), ..., we start two intervals later with (1, 2, 0, 6), (1, 4, 2, 0), ..., and proceed as before we obtain $\chi^2 = M/C = 17{\cdot}58$,

a value which, for $\nu = 12$, is not significant at the 10 % level. In other cases, which we have examined, shifting the position of the grouping divisions has not led to so large a change in χ^2. However, the limitation of the test must be recognized. In this case we cannot apply the test to single intervals and so avoid ambiguity (as was possible in examining the data in Table 1) because of the presence of zero intervals in the data as recorded.

Finally, we think it must be concluded that:

(*a*) there is definite evidence of differences in accident expectation between Divisions;

(*b*) the interval of 98 days in Division 4 is exceptional;

(*c*) apart from this we have not established inconsistency with the hypothesis that the intervals are randomly and exponentially distributed within a division.

8. Variations in the expectation

If E is the expectation of accidents per unit time and is a constant parameter, it has been shown that the distribution of time intervals between events is exponential and is given by the probability density function

$$f(t) = E\,e^{-Et}. \tag{1}$$

The distribution of accidents per unit time is the well known Poisson distribution which is discontinuous, the probability of r accidents in time τ being

$$p(r) = \frac{e^{-E\tau}(E\tau)^r}{r!}. \tag{11}$$

If E is itself a random variable with probability density function $h(E)$, $0 < E < \infty$, (1) may be transformed to give a new distribution of time intervals

$$f(t) = \int_{E=0}^{\infty} Eh(E)\,e^{-Et}\,dE. \tag{12}$$

This will be recognized as the Laplace transform. The tables of Laplace transforms and the considerable literature on their application to quite different problems may find some application to the analysis of interval distributions. See for example Carslaw & Jaeger (1941) and Doetsch (1947). Again, if E is a variable with probability distribution $h(E)$ as above, (11) may similarly be transformed to give the probability of intervals in time τ

$$p(r) = \frac{\tau^r}{r!}\int_0^{\infty} E^r h(E)\,e^{-E\tau}\,dE. \tag{13}$$

(13) is also a Laplace transform.

As an example, E may be assumed to be distributed in Pearson type III form

$$h(E) = \frac{c^q}{(q-1)!}E^{q-1}e^{-cE}. \tag{14}$$

Substitution in (12) and integration gives

$$f(t) = \frac{q}{c}\left(1+\frac{t}{c}\right)^{-(q+1)}. \tag{15}$$

This J-shaped curve is a form of the Pearson type XI distribution. Substitution of the Pearson type III distribution (14) in (13) gives the general term of the negative binomial distribution found by Greenwood & Yule (1920):

$$f(r) = \left(\frac{c}{c+1}\right)^q \frac{q(q+1)\ldots(q+r-1)}{r!\,(c+1)^r}, \tag{16}$$

where τ is taken to be unity. The Pearson type XI distribution is thus seen to be the distribution of the time intervals between events which have a negative binomial frequency distribution and a Pearson type III probability distribution of the expectation of events per unit time.

Every distribution of time intervals is seen to have two related distributions; the frequency distribution of events in fixed intervals of time, and the frequency distribution of the expectation of events per unit time. Their relationship is stated in general terms by (12) and (13).

9. CONCLUSIONS

It is apparent that many statistical tests, often already developed and tabulated for other purposes, may be usefully adapted to analyse industrial accident data if the difficulties of collecting and recording can be overcome.

None of the tests described in this paper demonstrates lack of homogeneity in the series of time intervals in Table 1. No significance attaches to the long intervals of 1613 and 1630 days or to the short intervals of 1 and 4 days. Such extremes might well occur during a period of 76 years even if the expectation of accident were constant. It is not suggested, however, that these tests are exhaustive or are the most sensitive that could be devised. Further, the conclusions only apply to one class of explosions in mines; milder explosions involving one man or more killed, or serious disasters involving fifty or more men killed may not provide homogeneous series. A thorough study of time intervals between explosions in mines would take into consideration the number of men killed and the number at risk.

It has been seen that the time intervals in Table 2 are not homogeneous. The range test, in particular, provides a method of identifying a time interval when the risk of accident was sufficiently low.

The problems of industrial accident control are statistically analogous to those of industrial quality control and similar practical techniques, including sequential procedures, can almost certainly be developed to help in their solution. There are, however, very important differences between quality control and accident control. It is easy to define categories of accident in which men are killed or in which men receive compensation, but many accidents are difficult to place in satisfactory categories. The problems of accidents are more human and more complicated than those of quality control. It is usually more important to extract the maximum amount of information from industrial accident data than from the data obtained from inspectors' samples. In quality control the size of sample can be chosen, but in accident control the data must be examined from a sample which may be any size and cannot be chosen by the investigator. Finally, accidents happen at a particular time and, as the time of occurrence is a valuable part of the data, techniques based on the analysis of time intervals, unfamiliar in quality control, are needed.

The analysis of intervals may not only be used to study events distributed in time, but points distributed in space. The exponential distribution was, indeed, probably first used to describe the distribution of free paths of molecules of a perfect gas. The analysis of intervals may be usefully applied to a much wider range of problems than hitherto.

Acknowledgement is made to the Ministry of Fuel and Power for permission to publish this paper.

REFERENCES

BARTLETT, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.
BORTKIEWICZ, L. (1898). *Bull. Inst. Statist.* **20** (2).
CARSLAW, H. S. & JAEGER, J. C. (1941). *Operational Methods in Applied Mathematics.* Oxford: Clarendon Press.
DOETSCH, G. (1947). *Tabellen zur Laplace Transformation und Anleitung zum Gebrauch.* Berlin: Springer.
FISHER, R. A. (1929). *Proc. Roy. Soc.* A, **125**, 54.
FISHER, R. A. (1950). *Contributions to Mathematical Statistics*, 16, 59*a*. London: Chapman and Hall.
GREENWOOD, M. (1946). *J.R. Statist. Soc.* **109**, 85.
GREENWOOD, M. & WOODS, H. M. (1919). *Industrial Fatigue Research Board Report*, 4. H.M.S.O.
GREENWOOD, M. & YULE, G. U. (1920). *J.R. Statist. Soc.* **83**, 255.
HARTLEY, H. O. (1951). *Biometrika*, **37**, 271.
JEFFREYS, H. (1948). *The Theory of Probability*, 2nd ed. Oxford University Press.
MORAN, P. A. P. (1951). *J.R. Statist. Soc.*, Series B, **13**, 147.
MORANT, G. M. (1920). *Biometrika*, **13**, 309.
NEYMAN, J. & PEARSON, E. S. (1928). *Biometrika*, **20A**, 175.
NEYMAN, J. & PEARSON, E. S. (1931). *Bull. Int. Acad. Cracovie*, A, 460.
SUKHATME, P. V. (1936). *Statist. Res. Mem.* **1**, 94.
THOMPSON, C. M. & MERRINGTON, M. (1946). *Biometrika*, **33**, 296.
WHITWORTH, W. A. (1901). *Choice and Chance*, 5th ed. Cambridge: Deighton Bell and Co.
YULE, G. U. (1922). *J.R. Statist. Soc.* **85**, 95.

Note added in proof. The publication of a recent paper by Moran (1951) has drawn our attention to the fact that in the discussion following a paper by Greenwood (1946), regarding industrial accidents, Bartlett suggested that known tests for homogeneity of variances could be used in the analysis of random time intervals. In his paper, Moran supposes that as an alternative to the exponential distribution of time intervals of equation (1), the probability density function has the Type III form

$$f(t) = (pE)^p \, e^{-pEt} t^{p-1}/\Gamma(p).$$

He then shows that in testing the hypothesis that $p = 1$, i.e. that equation (1) is true against this class of alternatives, the likelihood ratio criterion will be a function of the M of equation (9), with $n = 1$.

The departure from randomness which Moran contemplates is, of course, different from that which we have considered, i.e. a situation in which the accident expectation E may change from time to time, but the distribution of t remains exponential.

THE ESTIMATION OF DEATH-RATES FROM CAPTURE-MARK-RECAPTURE SAMPLING

BY P. A. P. MORAN

Institute of Statistics, University of Oxford

Leslie & Chitty, in a recent paper (1951), have discussed the problem of the estimation of death-rates from the results of capture-mark-recapture samplings of animal populations. Their paper constitutes an important advance but leaves open a number of theoretical problems, and the theory underlying some of the methods used is still obscure. The purpose of the present paper is to clarify the assumptions made and the logic of some of the methods.

The basic idea is as follows. At successive points of time, which we take to be equidistant, a sample of animals is captured, marked with some mark distinctive of that particular sampling, and released. At each sampling the number of animals showing each of the various possible combinations of previous marks is recorded. From these records it is possible, under various assumptions, to make an estimate of the death-rate. We suppose the samplings to take place at times $t_0, t_1, \ldots, t_i, \ldots$ which are equidistant. Let N_0 be the unknown total population at time t_0 and R_0 the size of the sample then taken. Each member of this sample is given the mark '0'. Having returned this sample to the population another sample of size R_1 is taken at time t_1. The number of this sample which possess the mark '0' is defined to be r_{01} and the rest as u_1, so that $R_1 = r_{01} + u_1$. Similarly, the sample at time t_2, of size R_2, is composed of r_{02} members with marks '0' and '2' but not '1', r_{12} with marks '1' and '2' but not '0', and of r_{012} with all three marks '0', '1' and '2'. Finally, there are u_2 individuals with no previous mark. Similarly, the sample of size R_n, taken at time t_n, is made up of u_n individuals not previously marked, together with various groups of members $r_{ab\ldots j}$ which have been previously marked at times $a, b, \ldots, j$, etc. For t_0, t_1, t_2 and t_3 the possible classes are set out in Table 1.

Table 1

t_0	t_1		t_2		t_3	
u_0	PR_0	r_{01}	Pr_{01}	r_{012}	Pr_{012}	r_{0123}
R_0	$N_1 - PR_0$	u_1	$P^2R_0 - Pr_{01}$	r_{02}	$P^2r_{01} - Pr_{012}$	r_{013}
		R_1	Pu_1	r_{12}	Pr_{02}	r_{023}
			$N_2 - P^2R_0 - Pu_1$	u_2	Pr_{12}	r_{123}
				R_2	$P^3u_0 - P^2r_{01} - Pr_{02}$	r_{03}
					$P^2u_1 - Pr_{12}$	r_{13}
					Pu_2	r_{23}
					$N_3 - P^3u_0 - P^2u_1 - Pu_2$	u_3
						R_3

We now assume that there is a constant death-rate acting during each period of time (t_i, t_{i+1}) so that a proportion $Q = 1 - P$ die off. Thus of the R_0 animals marked at time t_0 the expected number surviving at time t_1 is PR_0. Let N_1 be the total population at time t_1. We do not take this to be a random variable nor do we specify it in any way as we wish to allow for an arbitrary number of births (or immigration but not emigration) during the period (t_0, t_1). Similarly, at time t_2, we suppose that the expected number of animals in the population with marks '0' and '1' to be Pr_{01}, with mark '0' but not '1' to be $P^2R_0 - Pr_{01}$, with mark '1' but not '0' to be Pu_1, and without any mark to be $N_2 - P^2R_0 - Pu_1$, where N_2 is the total population at time t_2. Then in the sample of R_2 animals taken at time t_2 the observed values r_{012}, r_{02}, r_{12} and u_2 will have expectations proportional to these quantities. The corresponding sample values, with the expected numbers in the population at time t_3, are also given in Table 1, which can be easily extended to later values of t. We also notice that Table 1 gives the expected values in the population at each time *conditional on the sample values occurring in the previous samples*. These are the quantities we need if we wish to write down an expression for the probability of the whole observed series of samples.

We have still not properly defined the sampling model involved, and we here run into some difficulty. Leslie & Chitty distinguish 'deterministic' and 'probabilistic' models, but in their sampling experiment they use a model of a third type which we shall call the 'incomplete probabilistic model'. The distinction between the three models arises from the exact meaning given to the survival rate P.

The deterministic model

In this model P is taken to be the exact proportion of the population, in each class, surviving from one sampling time to the next. Then in order to write down the exact distribution at time t_1, say, PR_0 must be an integer. r_{01} is then distributed in a hypergeometric distribution, since the sample of R_1 is a random sample out of a total finite population of N_1 members, PR_0 of which are marked '0' and $N_1 - PR_0$ unmarked. Pr_{01} cannot then, with certainty, be asserted to be a whole number. An exact distribution theory for this model cannot therefore be set up unless $P = 1$. We must therefore approximate to the truth by ignoring this difficulty for the moment and supposing that PR_0, Pr_{01}, $P^2R_0 - Pr_{01}$, etc., can be regarded as integers. At each stage we then assume that we have random sampling from a finite population divided into several classes. If such a population consists of n members of which $n_1, \ldots, n_k$ belong to classes $1, \ldots, k$ $(\Sigma n_i = n)$, and we take a random sample of m $(m < n)$, the probability of obtaining $m_1, \ldots, m_k$ $(\Sigma m_i = m)$ in these classes is

$$\binom{n}{m}^{-1}\binom{n_1}{m_1}\ldots\binom{n_k}{m_k}. \tag{1}$$

We apply this result to the above scheme, noticing that the distribution of the kth sample is completely determined by what has already happened in samplings at $t_0, \ldots, t_{k-1}$. The probability to be attached to the first three samplings is therefore

$$\binom{N_1}{R_1}^{-1}\binom{PR_0}{r_{01}}\binom{N_1 - PR_0}{u_1}\binom{N_2}{R_2}^{-1}\binom{Pr_{01}}{r_{012}}\binom{P^2R_0 - Pr_{01}}{r_{02}}\binom{Pu_1}{r_{12}}\binom{N_2 - Pu_1 - P^2R_0}{u_2}. \tag{2}$$

This expression can easily be extended to further sampling stages. It clearly cannot represent an exact model of what is happening unless $P = 1$ (when $P \neq 1$ the above expressions may be

interpreted in terms of Beta functions). However, we can expect to obtain a good estimator for P in an actual situation by maximizing (2) for variations in P. This maximization is somewhat awkward, and Leslie & Chitty therefore approximate to (2) by replacing the multiple hypergeometric distributions by multinomial distributions. (2) then becomes

$$\binom{R_1}{r_{01}}\frac{(PR_0)^{r_{01}}(N_1-PR_0)^{u_1}}{N_1^{R_1}}\quad\frac{R_2!}{r_{012}!\,r_{02}!\,r_{12}!\,u_2!}\quad\frac{(Pr_{01})^{r_{012}}(P^2R_0-Pr_{01})^{r_{02}}(Pu_1)^{r_{12}}(N_2-Pu_1-P^2R_0)^{u_2}}{N_2^{R_2}} \tag{3}$$

This is now a genuine probability distribution whatever the value of P in the range $(0, 1)$, and we may therefore regard this as a probability distribution defining the 'deterministic model', although it cannot be other than an approximation to what really happens. If we now wish to estimate P, N_1 and N_2 we maximize (3) for variations in these quantities (the estimation of N_0 being, of course, impossible). In their paper Leslie & Chitty confine themselves to the estimation of P. To do this, they ignore the distribution of the u's and consider the conditional distribution, at each stage, of the r's, keeping the u's fixed. This is clearly valid, and the expression for the conditional likelihood which is to be maximized can then be written (for three sampling times)

$$\frac{(R_2-u_2)!}{r_{012}!\,r_{02}!\,r_{12}!}\quad\frac{(Pr_{01})^{r_{012}}(P^2R_0-Pr_{01})^{r_{02}}(Pu_1)^{r_{12}}}{(Pu_1+P^2R_0)^{R_2-u_2}}, \tag{4}$$

since the distribution at sampling time t_1 contributes no information about P. Omitting constant factors and cancelling P from the numerator and denominator of (4), the expression to be maximized is

$$\frac{(PR_0-r_{01})^{r_{02}}}{(u_1+PR_0)^{R_2-u_2}}.$$

The corresponding expressions for further stages can easily be written down. Thus for four stages we include the distribution at t_3 which, omitting a constant factor, is

$$\frac{(Pr_{01}-r_{012})^{r_{013}}(P^2u_0-Pr_{01}-r_{02})^{r_{03}}(Pu_1-r_{12})^{r_{13}}}{(P^2u_0+Pu_1+u_2)^{R_3-u_3}}. \tag{5}$$

This procedure does not, in fact, throw away any information about P (apart from that resulting from using such approximations as equation (3)) and Leslie has shown (results not yet published) that the joint estimation of P and the N's leads to the present method being used to estimate P first and the N's afterwards. Of course, if some restraint is placed on the N_i, as, for example, taking them to be all equal, or declining exponentially, some small, but probably very small, loss of information will result from the use of the present theory.

When more than three samplings are involved, the maximization of the likelihood becomes a fairly lengthy numerical job and various methods of grouping terms together have therefore been proposed. Such methods may throw away information, but do make the calculations easier. Before discussing them we consider the two other models.

The probabilistic model

We now suppose that P really is a probability of survival. Taking it to be the same for each animal, the population at sampling stage 1 may be written

$$R_{01},$$
$$U_1,$$

where R_{01} is the total number of animals in the population with the mark 0 and $U_1 = N_1 - R_{01}$ is the rest. R_{01} is a random variable distributed in a binomial distribution with total R_0 and expectation PR_0. Then, the sample taken at time t_1 is a random sample out of this population. A similar situation exists at later times. Thus in this model probability enters in two ways: first in the probability distribution of the survivors in each marked class, and second in the randomness of the sample taken at each time t_i. This set-up, although providing a real probability model and being much closer to what is actually happening in practice, results in a very much more complicated theory which I have not attempted to work out. However, it does seem intuitively clear that the variance of the estimator obtained will be larger than for the deterministic model. This suggests that in actual practice the use of the estimate of variance provided by the deterministic model will suggest a greater practical accuracy than is justified. Although the sampling experiments carried out by Leslie & Chitty use what is here called the 'semi-probabilistic' model the same point should hold. It is therefore somewhat curious that in actual fact their estimates of variance, based on the hypothesis of a deterministic model, are very close to the observed variance of the estimators in their sampling experiment.

The semi-probabilistic model

One of the difficulties in setting up the probabilistic model is that there is a non-zero, albeit very small, probability that so many animals die that there are less than R_i left at time t_i. R_i cannot therefore be prescribed before the experiment begins. Leslie & Chitty avoid this difficulty by altering the meaning of P. Instead of P representing a fixed proportion surviving in each class (as in the deterministic model), or a probability of each individual surviving, they suppose that it represents a fixed proportion of the total population which survives from t_i to t_{i+1}, but that the $N_i P$ individual animals which do survive are chosen at random without reference to their markings. Unless the total of marked animals in the population is both small and at the same time a large fraction of the total number, the actual numbers surviving in each marked class will have probability distributions which will be very closely approximated by those of the probabilistic model. The advantage of this scheme in an experiment with an urn and counters is that it is easily carried out and can be arranged so that at each time t_i there are always more than R_i animals in the population to be sampled. As pointed out above the introduction of this new element of randomness should result in our estimator of P having a variance which is inflated above that of the deterministic model. It is possible that the reason why this is not observed in Leslie & Chitty's experiments is that the scale of the experiment is not large enough for the estimator of variance to be really accurate.

Methods of grouping

We now return to the deterministic mode and suppose that the probability of the observed values is given by (4), multiplied by (5) if sampling at t_3 is included, and by further factors if necessary. Consider first the case where samplings at t_0, t_1 and t_2 only are taken. To estimate P the natural thing to do is to take the logarithm of (4) and maximize it with respect to variations in P. Ignoring the difficulties resulting from the approximate nature of the model, this method is clearly most efficient, at least asymptotically, but in practice the calculations quickly become lengthy if four or more samplings are used. It has therefore been suggested that before writing down the formulae for the likelihood, the various r's

should be grouped together and two such methods of grouping have been used. The first of these (method A), used by Jackson (1939, 1948) and Fisher & Ford (1947), is the simplest numerically but appears to throw away some information. Moreover, its logical basis is very obscure, and this obscurity was the principal reason for writing the present paper.

Table 2

$k_{01} = r_{01}$	$s'_1 = k_{01}$
$k_{02} = r_{012} + r_{02}$	$s'_2 = k_{02} + k_{12}$
$k_{12} = r_{012} + r_{12}$	
$k_{03} = r_{0123} + r_{013} + r_{023} + r_{03}$	
$k_{13} = r_{0123} + r_{013} + r_{123} + r_{13}$	$s'_3 = k_{03} + k_{13} + k_{23}$
$k_{23} = r_{0123} + r_{023} + r_{123} + r_{23}$	

Table 3

t_1	t_2	t_3
PR_0 k_{01}	P^2R_0 k_{02} PR_1 k_{12}	P^3R_0 k_{03} P^2R_1 k_{13} PR_2 k_{23}
PR_0 s'_1	$P^2R_0 + PR_1$ s'_2	$P^3R_0 + P^2R_1 + PR_2$ s'_3

Roughly speaking, the idea is to consider the distribution of the individual marks in the observed sample, rather than the distribution of the marked individuals. The r's are grouped into k's in the manner shown in Table 2, the u's are ignored, and we set up a joint distribution for the k's From this we see that k_{ij} is the number of animals in the sample taken at time t_j which have a mark put on at time t_i. The expected number in the population at time t_i with this mark is $P^{j-i}R_i$, so that we may now set out a scheme as in Table 3, following Leslie & Chitty (1951, p. 273). The s'_i are the total number of marks corresponding to $t_0, \ldots, t_{i-1}$ found on the animals at the sampling at time t_i. The simplicity of this table compared with Table 1 is one of the reasons for the attractiveness of method A. Leslie & Chitty then write for the log-likelihood (at time t_2 say) expressions such as

$$k_{02} \log P - s'_2 \log (PR_0 + R_1), \tag{6}$$

a factor P and a constant having been omitted. This is apparently the method used by Fisher & Ford and amounts to assuming that for constant s'_2, k_{02} is distributed binomially with probability

$$\frac{P^2R_0}{P^2R_0 + PR_1}. \tag{7}$$

We are thus trying to infer the value of P by referring the observed sample values to a hypothetical population of all samples at time t_2 which give the same total s'_2 of marks, and so we are dealing with a population of marks instead of a population of marked animals. It is nevertheless puzzling that we can treat k_{02} as a binomial variate, even approximately, when we notice that k_{02} and $k_{12} = s'_2 - k_{02}$ have the quantity r_{012} in common.

To see what is happening we return to a simpler case. We suppose $P = 1$ and $N_0 = N_1 = N_2$. We can then take the joint distribution of r_{01}, r_{012}, r_{02} and r_{12} to be given by the distribution (2) which is exact when $P = 1$. It will now be shown that if we first average over all possible values of k_{01}, keeping k_{02} and k_{12} fixed and then consider the joint distribution of k_{02} and k_{12} (i.e. we consider the whole distribution at times t_1 and t_2 and not just t_2 alone) then they are independently distributed in hypergeometric distributions with the expectations based on Table 3 and with $P = 1$. We shall also see that this result is, in fact, almost obvious intuitively.

Now if k_{02} and k_{12} are thus independently distributed and if their expectation is fairly small compared with R_2, they can be regarded as approximately distributed in Poisson distributions. Assuming this, the conditional distribution of one of them, given that the sum of the two is fixed, is a binomial distribution as assumed in the expression for likelihood (6). For suppose x and y are variates independently distributed in Poisson distributions with means λ and μ. The joint probability of x and y is

$$e^{-\lambda-\mu}\frac{\lambda^x\mu^y}{x!\,y!},$$

and the conditional distribution of x, given that $x+y = N$, is

$$e^{-\lambda-\mu}\frac{\lambda^x\mu^y}{x!\,y!}\left\{e^{-\lambda-\mu}\sum_{s=0}^{s=N}\frac{\lambda^s\mu^{N-s}}{s!\,N-s!}\right\}^{-1} = \binom{N}{s}\left(\frac{\lambda}{\lambda+\mu}\right)^x\left(\frac{\mu}{\lambda+\mu}\right)^{N-x}.$$

The Poisson distribution is the only distribution for which this is true.

In fact it can be proved (Moran, 1952) that if x and y are two independent variates with distributions on the non-negative integers, and if the distribution of x given that of $x+y$ is binomial, then, with the exception of a trivial case, x and y must each be distributed in Poisson distributions, so that the Poisson distribution is a necessary condition for the above formula to hold exactly.

Now taking $P = 1$ and $N_0 = N_1 = N_2 = N$, say, the exact probability of the observed values r_{01}, r_{012}, r_{02} and r_{12} is

$$\begin{aligned}\binom{N}{R_1}^{-1}\binom{N}{R_2}^{-1}\binom{R_0}{r_{01}}\binom{N-R_0}{R_1-r_{01}}\binom{R_0-r_{01}}{r_{02}}\binom{R_1-r_{01}}{r_{12}}\binom{r_{01}}{r_{012}}\binom{N-R_0-R_1+r_{01}}{u_2}\\ = \binom{N}{R_1}^{-1}\binom{N}{R_2}^{-1}\left\{\frac{R_0!\,N-R_0!}{r_{02}!\,r_{12}!\,r_{012}!\,R_2-r_{02}-r_{12}-r_{012}!}\right\}A, \qquad (8)\end{aligned}$$

where $$A = \{R_0-r_{01}-r_{02}!\,R_1-r_{01}-r_{12}!\,r_{01}-r_{012}!\,N-R_0-R_1+r_{01}-u_2!\}^{-1}.$$

We now have to sum over all values of r_{01}, keeping r_{02}, r_{12} and r_{012} fixed. It is easy to see that the range of permissible values for r_{01} is given by

$$r_{012} \leqslant r_{01} \leqslant \min(R_1-r_{12}, R_0-r_{02}) = m \text{ (say)},$$

both these limits being attainable. Summing over these values we have

$$\begin{aligned}\Sigma A &= \{R_1-r_{012}-r_{12}!\,N-R_1-R_2+r_{012}+r_{12}!\}^{-1}\,\Sigma\binom{R_1-r_{012}-r_{12}}{r_{01}-r_{012}}\binom{N-R_1-R_2+r_{012}+r_{12}}{R_0-r_{01}-r_{02}}\\ &= \{R_1-k_{12}!\,N-R_1-R_2+k_{12}!\}^{-1}\,\Sigma\binom{R_1-k_{12}}{r_{01}-r_{012}}\binom{N-R_1-R_2+k_{12}}{R_0-r_{01}-r_{02}}\\ &= \{R_1-k_{12}!\,N-R_1-R_2+k_{12}!\}^{-1}\sum_{s=0}^{m}\binom{R_1-k_{12}}{s}\binom{N-R_1-R_2+k_{12}}{R_0-k_{02}-s}\\ &= \{R_1-k_{12}!\,N-R_1-R_2+k_{12}!\}^{-1}\end{aligned}$$

multiplied by the coefficient of $x^{R_0-k_{02}}$ in

$$(1+x)^{R_1-k_{12}}(1+x)^{N-R_1-R_2+k_{12}} = (1+x)^{N-R_2}.$$

This coefficient is $\binom{N-R_2}{R_0-k_{02}}$.

The joint probability of the observed values at the third sampling, when averaged over all values at the second, can therefore be written

$$\binom{N}{R_1}^{-1}\binom{N}{R_2}^{-1} \times \left\{\frac{R_0!\,N-R_0!\,N-R_2!}{r_{02}!\,r_{12}!\,r_{012}!\,R_2-r_{02}-r_{12}-r_{012}!\,R_0-k_{02}!\,N-R_0-R_2+k_{02}!\,R_1-k_{12}!\,N-R_1-R_2+k_{12}!}\right\}. \tag{9}$$

We now want to keep k_{02} and k_{12} fixed and sum (9) over all possible values of r_{012}, which can clearly take all values between 0 and min (k_{02}, k_{12}) inclusive. Consider therefore the sum

$$\sum_{r_{012}=0}^{\min(k_{02},k_{12})}\{r_{02}!\,r_{12}!\,r_{012}!\,R_2-r_{02}-r_{12}-r_{012}!\}^{-1} = \{k_{02}!\,R_2-k_{02}!\}^{-1}\sum_{r_{012}=0}^{\min(k_{02},k_{12})}\binom{k_{02}}{r_{012}}\binom{R_2-k_{02}}{k_{12}-r_{012}}.$$

The sum is the coefficient of $x^{k_{12}}$ in the expression

$$(1+x)^{k_{02}}(1+x)^{R_2-k_{02}},$$

and so the above is equal to $(k_{02}!\,R_2-k_{02}!)^{-1}\binom{R_2}{k_{12}}$.

Inserting this in (9) we get

$$\binom{N}{R_1}^{-1}\binom{N}{R_2}^{-1}\left\{\frac{R_0!\,N-R_0!\,R_2!\,N-R_2!}{R_0-k_{02}!\,N-R_0-R_2+k_{02}!\,R_1-k_{12}!\times N-R_1-R_2+k_{12}!\,R_2-k_{12}!\,k_{12}!\,k_{02}!\,R_2-k_{02}!}\right\} \tag{10}$$

$$= \binom{N}{R_2}^{-2}\left\{\frac{R_0!\,N-R_0!\,R_1!\,N-R_1!}{R_0-k_{02}!\,N-R_0-R_2+k_{02}!\,R_1-k_{12}!\,N-R_1-R_2+k_{12}!\,R_2-k_{12}!\,k_{12}!\,k_{02}!\,R_2-k_{02}!}\right\}$$

$$= \binom{N}{R_2}^{-2}\binom{R_0}{k_{02}}\binom{N-R_0}{R_2-k_{02}}\binom{R_1}{k_{12}}\binom{N-R_1}{R_2-k_{12}}, \tag{11}$$

which shows that k_{02} and k_{12} are independently distributed in hypergeometric distributions. It must be noticed that this result has been reached by averaging out the results of the sampling at t_1 and that we are not here dealing with a conditional distribution at time t_2 given that at time t_1. The result expressed by (11) can in fact be seen intuitively. For the process of three successive samplings of the same N animals can be regarded as three independent actions of the experimenter in each of which a certain mark is put on numbers R_0, R_1 and R_2 of the animals selected at random.

Consider those animals which were selected and marked at time t_2. At times t_0 and t_1, marks '0' and '1' are put on R_0 and R_1 animals, respectively, in an independent and random manner. Then k_{02}, which is the number of animals in the sample at time t_2 which possess the mark '0', and may or may not possess the mark '1', is clearly distributed in a hypergeometric distribution which is independent of the similar distribution of k_{12}.

It is difficult to see how to give an exact mathematical discussion of the case where $P \neq 1$. Presumably one might take (3) and sum out r_{01} and then r_{012}, but this seems to be very difficult mathematically. However, it seems intuitively clear that k_{02} and k_{12} will still be independent.

We can clearly extend the above type of argument to discuss the joint distribution of k_{03}, k_{13} and k_{23}. Here we meet another and more important difficulty in this type of grouping. For (taking the case $P = 1$) the joint distribution of k_{03}, k_{13} and k_{23} can be shown to consist of the product of three independent hypergeometric distributions, *when* r_{01}, r_{012}, r_{02} *and* r_{12} *are averaged out*. This means that even if we are justified in using a multinomial distribution for k_{03}, k_{13} and k_{23} (as an approximation) we are not considering a distribution at time t_3 conditional on what has already happened, but are averaging over all possible previous results. But as can be seen from the discussion of method A given by Leslie & Chitty, this method involves writing down the part of the likelihood at time t_3 as if it did not depend on the previous results and so we are not using the true likelihood of the whole experiment. This introduces a further approximation and possibly a bias of unknown extent into method A.

Thus we may sum up as follows: (*a*) For reasons of mathematical convenience it appears necessary to use a model which does not take account of some randomness in the situation and therefore estimates of variance are likely to be less than the true values by an unknown amount. (*b*) Methods of grouping can be used to cut down the computational labour but of these the method used by Jackson and by Fisher & Ford throws away information and has other theoretical objections. The method used by Leslie & Chitty is fully efficient, under the assumption of the deterministic model, and appears to be well worth the extra labour required.

REFERENCES

Fisher, R. A. & Ford, E. B. (1947). *Heredity*, **1**, 143.
Jackson, C. H. N. (1939). *J. Anim. Ecol.* **8**, 238.
Jackson, C. H. N. (1948). *Ann. Eugen., Lond.*, **14**, 91.
Leslie, P. H. & Chitty, D. H. (1951). *Biometrika*, **38**, 269.
Moran, P. A. P. (1952). *Proc. Camb. Phil. Soc.* (in the Press).

MISCELLANEA

A note on the design problem

By K. D. TOCHER, *Imperial College, London*

If an experiment gives rise to a series of uncorrelated observations of equal variance σ^2, denoted by the vector y, and we have $\mathscr{E}(y) = a\theta$, where a is a matrix of known constants and full rank, and θ is a vector of unknown parameters, then Markoff's theorem gives the minimum variance unbiased estimates of the parameter θ as $\hat{\theta} = (a'a)^{-1}a'y$ and $\mathscr{V}(\hat{\theta}) = (a'a)^{-1}\sigma^2$.

The central problem of design in experimentation is the choice of the elements of a to minimize the diagonal elements of $(a'a)^{-1}$. The well-known solution of this problem is that a shall have orthogonal columns. It may be of some interest to give a short and simple proof of this which is independent of Hadamard's theorem or the ideas of correlation used in previous proofs.

The proof hinges on the triangular resolution of a positive definite symmetric matrix (Turing, 1948). This solution forms the basis of several methods of inverting matrices rapidly gaining popularity, but its value as a technique in theoretical problems concerning matrices does not seem to be sufficiently realized.

It is clear that if the elements of a are allowed to increase indefinitely, then the elements of $(a'a)^{-1}$ will decrease indefinitely, and so some restraint on the elements of a must be imposed to ensure a realistic solution. We shall follow customary practice and assume that this restraint fixes the sum of squares of the ith column of a as c_i.

Let u be an upper triangular matrix ($u_{ij} = 0$ if $i > j$) such that $(a'a) = u'u$.

Since $(a'a)$ is positive definite, u is non-singular and has an inverse v which is upper triangular. Clearly $v_{ii} = 1/u_{ii}$ and $(a'a)^{-1} = vv'$. Suppose the ith diagonal element of $(a'a)^{-1}$ is α_i, then

$$\alpha_i = \sum_j v_{ij}^2 \geqslant v_{ii}^2 = 1/u_{ii}^2 \geqslant 1/\sum_j u_{ji}^2 = 1/c_i.$$

For α_i to achieve its lower bound $1/c_i$ both inequalities must degenerate to equalities, and for this to occur, for all i, all the elements v_{ij} with $i \neq j$ must vanish. This will establish the first in each pair of equalities required. In that case v will be diagonal, and hence so will u, showing that the second of each pair of equalities also will be satisfied.

Thus the condition to achieve the lower bounds simultaneously is that $a'a$ is diagonal, or equivalently, that a has orthogonal columns.

REFERENCE

Turing, A. M. (1948). *Quart. J. Mech. Appl. Math.* **1**, 287.

Tables of percentage points of the 'Studentized' extreme deviate from the sample mean

By K. R. NAIR

Forest Research Institute, Dehra Dun, India

Denote by $x_{(1)} \ldots x_{(n)}$ a random sample of n observations drawn from a normal population with standard deviation σ. Let $x_1 \ldots x_n$ be the same sample arranged in ascending order of magnitude so that x_r is the rth ranked (or ordered) variate in the sample $\{x_{(i)}\}$.

If an estimate s_ν of the unknown σ is available with ν degrees of freedom *independent* of the sample $\{x_{(i)}\}$ the present author suggested the use of $(x_n - \bar{x})/s_\nu$ or $(\bar{x} - x_1)/s_\nu$ as a test criterion for a single outlier x_n (or x_1). In Tables 6 A and B of the author's (1948) paper the lower and upper 5 and 1 % points of the distribution of the 'Studentized' extreme deviate, $(x_n - \bar{x})/s_\nu$ or $(\bar{x} - x_1)/s_\nu$ were obtained for $n = 3$ to 9 and for selected values of $\nu \geqslant 10$. Those tables have now been extended to cover a few more per cent points, namely, 10, 2·5, 0·5 and 0·1. The values for all the six points, including the 5 and 1 % points previously published, are given in Tables 1 A and B below.

The method used for constructing these tables and examples illustrating the use of the 'Studentized' extreme deviate in the analysis and interpretation of designed experiments have been given in the author's (1948) paper to which reference is invited.

Table 1 A. *Lower percentage points of the 'Studentized' extreme deviate* $(x_n - \bar{x})/s_\nu$ *or* $(\bar{x} - x_1)/s_\nu$.

n / ν	3	4	5	6	7	8	9	3	4	5	6	7	8	9
	10 % points							5 % points						
10	0·29	0·45	0·57	0·67	0·74	0·81	0·87	0·20	0·35	0·46	0·55	0·62	0·69	0·74
15	0·29	0·45	0·57	0·67	0·75	0·82	0·88	0·20	0·35	0·46	0·55	0·63	0·70	0·75
30	0·29	0·45	0·58	0·67	0·76	0·83	0·89	0·20	0·35	0·46	0·56	0·64	0·70	0·77
∞	0·29	0·45	0·58	0·68	0·76	0·84	0·90	0·20	0·35	0·47	0·56	0·65	0·72	0·78
	2·5 % points							1 % points						
10	0·14	0·27	0·37	0·46	0·53	0·59	0·64	0·09	0·19	0·29	0·37	0·43	0·49	0·54
15	0·14	0·27	0·37	0·46	0·54	0·60	0·66	0·09	0·19	0·29	0·37	0·44	0·50	0·56
30	0·14	0·27	0·38	0·47	0·55	0·61	0·67	0·09	0·20	0·29	0·38	0·45	0·51	0·57
∞	0·14	0·27	0·38	0·48	0·56	0·62	0·69	0·09	0·20	0·30	0·38	0·46	0·53	0·59
	0·5 % points							0·1 % points						
10	0·06	0·15	0·24	0·31	0·38	0·43	0·48	0·03	0·09	0·16	0·22	0·28	0·33	0·37
15	0·06	0·15	0·24	0·32	0·38	0·44	0·50	0·03	0·09	0·16	0·22	0·28	0·34	0·39
30	0·06	0·15	0·24	0·32	0·39	0·45	0·51	0·03	0·09	0·16	0·23	0·29	0·35	0·40
∞	0·06	0·16	0·25	0·33	0·40	0·47	0·53	0·03	0·09	0·16	0·23	0·30	0·36	0·41

Table 1 B. *Upper percentage points of the 'Studentized' extreme deviate* $(x_n - \bar{x})/s_\nu$ *or* $(\bar{x} - x_1)/s_\nu$.

n / ν	10 % points							5 % points						
	3	4	5	6	7	8	9	3	4	5	6	7	8	9
10	1·68	1·93	2·11	2·25	2·36	2·46	2·54	2·02	2·29	2·49	2·63	2·75	2·85	2·93
11	1·66	1·91	2·08	2·21	2·32	2·42	2·49	1·99	2·26	2·44	2·58	2·70	2·79	2·87
12	1·65	1·89	2·05	2·19	2·29	2·38	2·46	1·97	2·22	2·40	2·54	2·65	2·75	2·83
13	1·63	1·87	2·04	2·16	2·27	2·36	2·43	1·95	2·20	2·38	2·51	2·62	2·71	2·79
14	1·62	1·85	2·02	2·14	2·25	2·33	2·41	1·93	2·18	2·35	2·48	2·59	2·68	2·76
15	1·61	1·84	2·00	2·13	2·23	2·31	2·39	1·92	2·16	2·33	2·46	2·56	2·65	2·73
16	1·61	1·83	1·99	2·12	2·22	2·30	2·37	1·90	2·14	2·31	2·44	2·54	2·63	2·70
17	1·60	1·82	1·98	2·10	2·20	2·28	2·35	1·89	2·13	2·30	2·42	2·52	2·61	2·68
18	1·59	1·82	1·97	2·09	2·19	2·27	2·34	1·88	2·12	2·28	2·41	2·51	2·59	2·66
19	1·59	1·81	1·97	2·09	2·18	2·26	2·33	1·87	2·11	2·27	2·39	2·49	2·58	2·65
20	1·58	1·80	1·96	2·08	2·17	2·25	2·32	1·87	2·10	2·26	2·38	2·48	2·56	2·63
24	1·57	1·78	1·94	2·05	2·15	2·22	2·29	1·84	2·07	2·23	2·35	2·44	2·52	2·59
30	1·55	1·77	1·92	2·03	2·12	2·20	2·26	1·82	2·04	2·20	2·31	2·40	2·48	2·55
40	1·54	1·75	1·89	2·01	2·09	2·17	2·23	1·80	2·02	2·17	2·28	2·37	2·44	2·51
60	1·52	1·73	1·87	1·98	2·07	2·14	2·20	1·78	1·99	2·14	2·25	2·33	2·41	2·47
120	1·51	1·71	1·85	1·96	2·05	2·12	2·18	1·76	1·97	2·11	2·21	2·30	2·37	2·43
∞	1·50	1·70	1·83	1·94	2·02	2·09	2·15	1·74	1·94	2·08	2·18	2·27	2·33	2·39

ν \ n	2·5 % points							1 % points						
	3	4	5	6	7	8	9	3	4	5	6	7	8	9
10	2·36	2·64	2·84	2·99	3·10	3·20	3·28	2·76	3·05	3·25	3·39	3·50	3·59	3·67
11	2·31	2·59	2·78	2·93	3·04	3·14	3·22	2·71	3·00	3·19	3·33	3·44	3·53	3·61
12	2·28	2·55	2·74	2·88	2·99	3·08	3·16	2·67	2·95	3·14	3·28	3·39	3·48	3·55
13	2·25	2·52	2·70	2·84	2·95	3·04	3·12	2·63	2·91	3·10	3·24	3·34	3·43	3·51
14	2·23	2·49	2·67	2·80	2·91	3·00	3·08	2·60	2·87	3·06	3·20	3·30	3·39	3·47
15	2·20	2·46	2·64	2·77	2·88	2·97	3·04	2·57	2·84	3·02	3·16	3·27	3·35	3·43
16	2·19	2·44	2·62	2·75	2·85	2·94	3·01	2·55	2·81	3·00	3·13	3·24	3·32	3·39
17	2·17	2·42	2·60	2·73	2·83	2·92	2·99	2·52	2·79	2·97	3·10	3·21	3·29	3·36
18	2·16	2·41	2·58	2·71	2·81	2·89	2·97	2·50	2·77	2·95	3·08	3·18	3·27	3·34
19	2·15	2·39	2·56	2·69	2·79	2·87	2·95	2·49	2·75	2·92	3·06	3·16	3·24	3·31
20	2·14	2·38	2·55	2·67	2·77	2·86	2·93	2·47	2·73	2·91	3·04	3·14	3·22	3·29
24	2·10	2·34	2·50	2·62	2·72	2·80	2·87	2·43	2·68	2·85	2·97	3·07	3·15	3·22
30	2·07	2·30	2·46	2·58	2·67	2·75	2·82	2·38	2·62	2·79	2·91	3·01	3·08	3·15
40	2·04	2·27	2·42	2·53	2·62	2·70	2·76	2·34	2·57	2·73	2·85	2·94	3·02	3·08
60	2·01	2·23	2·38	2·49	2·58	2·65	2·71	2·30	2·52	2·68	2·79	2·88	2·95	3·01
120	1·98	2·20	2·34	2·45	2·53	2·60	2·66	2·25	2·48	2·62	2·73	2·82	2·89	2·95
∞	1·95	2·16	2·30	2·41	2·49	2·56	2·61	2·22	2·43	2·57	2·68	2·76	2·83	2·88

ν \ n	0·5 % points							0·1 % points						
	3	4	5	6	7	8	9	3	4	5	6	7	8	9
10	3·03	3·32	3·52	3·65	3·76	3·85	3·92	3·54	3·84	4·04	4·17	4·28	4·35	4·40
11	2·98	3·27	3·46	3·60	3·71	3·79	3·86	3·49	3·80	3·99	4·12	4·23	4·30	4·36
12	2·93	3·22	3·41	3·55	3·66	3·74	3·81	3·45	3·75	3·94	4·07	4·19	4·26	4·31
13	2·89	3·18	3·37	3·50	3·61	3·70	3·77	3·41	3·71	3·90	4·03	4·14	4·22	4·28
14	2·86	3·14	3·33	3·46	3·57	3·66	3·73	3·38	3·67	3·86	4·00	4·10	4·18	4·24
15	2·83	3·11	3·29	3·43	3·53	3·62	3·69	3·35	3·64	3·83	3·96	4·06	4·15	4·21
16	2·80	3·08	3·26	3·39	3·50	3·58	3·66	3·32	3·51	3·80	3·93	4·03	4·12	4·18
17	2·77	3·05	3·23	3·36	3·47	3·55	3·62	3·29	3·58	3·77	3·90	4·00	4·09	4·15
18	2·75	3·02	3·21	3·34	3·44	3·53	3·60	3·27	3·55	3·74	3·88	3·98	4·06	4·12
19	2·73	3·00	3·18	3·31	3·42	3·50	3·57	3·25	3·53	3·72	3·85	3·95	4·03	4·10
20	2·72	2·98	3·16	3·29	3·39	3·48	3·55	3·23	3·51	3·70	3·83	3·93	4·01	4·08
24	2·66	2·92	3·09	3·22	3·32	3·40	3·47	3·16	3·44	3·62	3·75	3·85	3·93	4·00
30	2·60	2·86	3·02	3·15	3·25	3·32	3·39	3·08	3·36	3·53	3·66	3·76	3·84	3·90
40	2·55	2·79	2·96	3·08	3·17	3·25	3·31	3·01	3·27	3·44	3·57	3·66	3·74	3·81
60	2·50	2·73	2·89	3·01	3·10	3·17	3·23	2·93	3·19	3·35	3·47	3·56	3·64	3·70
120	2·45	2·67	2·83	2·94	3·02	3·09	3·15	2·85	3·10	3·26	3·37	3·46	3·53	3·59
∞	2·40	2·62	2·76	2·87	2·95	3·02	3·07	2·78	3·01	3·17	3·28	3·36	3·43	3·48

REFERENCE

NAIR, K. R. (1948). *Biometrika*, **35**, 118.

Extended and corrected tables of the upper percentage points of the 'Studentized' range

COMPUTED BY JOYCE M. MAY

EDITORIAL

1. *Need for revised tables.* Denote by $x_1, x_2, \ldots, x_n$ a random sample of n observations drawn from a normal population with standard deviation σ and arranged in ascending order of magnitude; then the range in the sample is given by $w = x_n - x_1$. Denote further by s^2 a 'mean-square' estimate of σ^2, independent of w, and based on ν degrees of freedom so that $\nu s^2/\sigma^2$ follows the χ^2 distribution for ν degrees of freedom. Then the *studentized range* is defined by

$$q = w/s = (x_n - x_1)/s. \qquad (1)$$

When, during the war, tables of the upper percentage points of q were computed (Pearson & Hartley, 1943) from those of the probability integral of w (Pearson & Hartley, 1942), an expansion of the 'Studentized' integral in powers of ν^{-1} was used (Hartley, 1944). Since it was realized that for small ν and large q this expansion would break down, these tables were not extended to values of ν below 10, and for $\nu \geqslant 10$ percentage points with values exceeding 6 were not regarded as reliable (Pearson & Hartley, 1943, p. 90, lines 25–6). Recently, Prof. J. W. Tukey has drawn attention to the possibility of a breakdown in the expansion at the larger values of q, resulting in errors of a magnitude far exceeding that anticipated by these warnings. This query has prompted us to arrange for a recomputation of the upper percentage points of q by the considerably more laborious but numerically exact method of quadrature described below, which, as will be seen from the attached tables, confirms Prof. Tukey's suspicions. In view of the more recent applications of 'Studentized' range (Patnaik, 1950; Hartley, 1950; David, 1951), it was felt that the recomputation of the more important upper percentage points should cover all small ν down to $\nu = 1$. For these very small values of ν an expansion derived by Pillai, and described elsewhere in this issue, has been found useful as a check. The upper 5 and 1 % points computed by Pillai for the range $1 \leqslant \nu \leqslant 4$, $2 \leqslant n \leqslant 8$ were found to agree with the tabled results of our computations to within a unit of the last figure quoted by Pillai.

Table of the upper percentage points of the 'Studentized' range $q = (x_n - x_1)/s$ (5 % *points*)

(n = sample size for range; ν = degrees of freedom of s)

ν \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18·0	26·7	32·8	37·2	40·5	43·1	45·4	47·3	49·1	50·6	51·9	53·2	54·3	55·4	56·3	57·2	58·0	58·8	59·6
2	6·09	8·28	9·80	10·89	11·73	12·43	13·03	13·54	13·99	14·39	14·75	15·08	15·38	15·65	15·91	16·14	16·36	16·57	16·77
3	4·50	5·88	6·83	7·51	8·04	8·47	8·85	9·18	9·46	9·72	9·95	10·16	10·35	10·52	10·69	10·84	10·98	11·12	11·24
4	3·93	5·00	5·76	6·31	6·73	7·06	7·35	7·60	7·83	8·03	8·21	8·37	8·52	8·67	8·80	8·92	9·03	9·14	9·24
5	3·61	4·54	5·18	5·64	5·99	6·28	6·52	6·74	6·93	7·10	7·25	7·39	7·52	7·64	7·75	7·86	7·95	8·04	8·13
6	3·46	4·34	4·90	5·31	5·63	5·89	6·12	6·32	6·49	6·65	6·79	6·92	7·04	7·14	7·24	7·34	7·43	7·51	7·59
7	3·34	4·16	4·68	5·06	5·35	5·59	5·80	5·99	6·15	6·29	6·42	6·54	6·65	6·75	6·84	6·93	7·01	7·08	7·16
8	3·26	4·04	4·53	4·89	5·17	5·40	5·60	5·77	5·92	6·05	6·18	6·29	6·39	6·48	6·57	6·65	6·73	6·80	6·87
9	3·20	3·95	4·42	4·76	5·02	5·24	5·43	5·60	5·74	5·87	5·98	6·09	6·19	6·28	6·36	6·44	6·51	6·58	6·65
10	3·15	3·88	4·33	4·66	4·91	5·12	5·30	5·46	5·60	5·72	5·83	5·93	6·03	6·12	6·20	6·27	6·34	6·41	6·47
11	3·11	3·82	4·26	4·58	4·82	5·03	5·20	5·35	5·49	5·61	5·71	5·81	5·90	5·98	6·06	6·14	6·20	6·27	6·33
12	3·08	3·77	4·20	4·51	4·75	4·95	5·12	5·27	5·40	5·51	5·61	5·71	5·80	5·88	5·95	6·02	6·09	6·15	6·21
13	3·06	3·73	4·15	4·46	4·69	4·88	5·05	5·19	5·32	5·43	5·53	5·63	5·71	5·79	5·86	5·93	6·00	6·06	6·11
14	3·03	3·70	4·11	4·41	4·64	4·83	4·99	5·13	5·25	5·36	5·46	5·56	5·64	5·72	5·79	5·86	5·92	5·98	6·03
15	3·01	3·67	4·08	4·37	4·59	4·78	4·94	5·08	5·20	5·31	5·40	5·49	5·57	5·65	5·72	5·79	5·85	5·91	5·96
16	3·00	3·65	4·05	4·34	4·56	4·74	4·90	5·03	5·15	5·26	5·35	5·44	5·52	5·59	5·66	5·73	5·79	5·84	5·90
17	2·98	3·62	4·02	4·31	4·52	4·70	4·86	4·99	5·11	5·21	5·31	5·39	5·47	5·55	5·61	5·68	5·74	5·79	5·84
18	2·97	3·61	4·00	4·28	4·49	4·67	4·83	4·96	5·07	5·17	5·27	5·35	5·43	5·50	5·57	5·63	5·69	5·74	5·79
19	2·96	3·59	3·98	4·26	4·47	4·64	4·79	4·92	5·04	5·14	5·23	5·32	5·39	5·46	5·53	5·59	5·65	5·70	5·75
20	2·95	3·58	3·96	4·24	4·45	4·62	4·77	4·90	5·01	5·11	5·20	5·28	5·36	5·43	5·50	5·56	5·61	5·66	5·71
24	2·92	3·53	3·90	4·17	4·37	4·54	4·68	4·81	4·92	5·01	5·10	5·18	5·25	5·32	5·38	5·44	5·50	5·55	5·59
30	2·89	3·48	3·84	4·11	4·30	4·46	4·60	4·72	4·83	4·92	5·00	5·08	5·15	5·21	5·27	5·33	5·38	5·43	5·48
40	2·86	3·44	3·79	4·04	4·23	4·39	4·52	4·63	4·74	4·82	4·90	4·98	5·05	5·11	5·17	5·22	5·27	5·32	5·36
60	2·83	3·40	3·74	3·98	4·16	4·31	4·44	4·55	4·65	4·73	4·81	4·88	4·94	5·00	5·06	5·11	5·15	5·20	5·24
120	2·80	3·36	3·69	3·92	4·10	4·24	4·36	4·47	4·56	4·64	4·71	4·78	4·84	4·90	4·95	5·00	5·04	5·09	5·13
∞	2·77	3·32	3·63	3·86	4·03	4·17	4·29	4·39	4·47	4·55	4·62	4·68	4·74	4·80	4·84	4·89	4·93	4·97	5·01

2. *Method of computation.* The computation of the present tables was based on the following expression for the probability integral of 'Studentized' range at the upper tail:

$$\Pr\{w/s \geqslant q\} = c_\nu \int_0^\infty [4sZ(s)]^\nu \frac{1}{s} p(qs \mid n)\, ds = c_\nu \int_0^\infty \left[4\frac{w}{q} Z\left(\frac{w}{q}\right)\right]^\nu \times \frac{1}{w}\, p(w \mid n)\, dw, \qquad (2)$$

where $Z(s) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}s^2}$, $c_\nu = 2\Gamma^{-1}(\frac{1}{2}\nu)\{\frac{1}{4}\sqrt{(\pi\nu)}\}^\nu$ and $p(w \mid n)$ is the upper tail probability integral of range in samples of n. The factor 4 has been included within the brackets, [ ], for the computational convenience of bringing the maximum of the function inside these brackets near to unity. A five-decimal MS. of $p(w \mid n)$ (of which 4 decimals were published in the form $1-p(w \mid n)$ by Pearson & Hartley (1942)) formed the basis for the master table of $w^{-1}p(w \mid n)$ which was computed for $n = 2\,(2)\,20$, $w = 0{\cdot}25\,(0{\cdot}25)\,2{\cdot}0\,(0{\cdot}5)\,8{\cdot}0$.

To form the integrand in (2) the other factor, viz. $\left[4\frac{w}{q} Z\left(\frac{w}{q}\right)\right]^\nu$, which depends on ν and q, was formed for $\nu = 1\,(1)\,4, 6, 8, 12$ and 24 and for some 6–8 equidistant values of $1/q$, covering the range of the upper 5 and 1% points of q which were estimated roughly for this purpose. For each combination of these values of n, ν and q, the integral in (2) was evaluated by numerical quadrature, using mainly Gregory's formula and utilizing, for small ν, any symmetry or antisymmetry at $w = 0$ in the integrand factors. For each combination of n and ν (forming, as it were, a 'master grid' in n and ν) the resulting integrals were differenced with regard to $1/q$ and the upper 5 and 1% point of q found by inverse interpolation. These values were checked by differencing both n-wise and ν-wise, and the remaining percentage points in the present table found by Lagrangian and/or harmonic interpolation in the values of the master grid. The resulting tables should be correct to within 3 units of the last figure quoted.

REFERENCES

David, H. A. (1951). *Biometrika*, **38**, 393.
Hartley, H. O. (1944). *Biometrika*, **33**, 173.
Hartley, H. O. (1950). *Biometrika*, **37**, 271.
Patnaik, P. B. (1950). *Biometrika*, **37**, 78.
Pearson, E. S. & Hartley, H. O. (1942). *Biometrika*, **32**, 301.
Pearson, E. S. & Hartley, H. O. (1943). *Biometrika*, **33**, 89.

Table of the upper percentage points of the 'Studentized' range $q = (x_n - x_1)/s$ (1% *points*)

(n = sample size for range, ν = degrees of freedom of s)

ν \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	90·0	134	164	186	202	216	227	237	246	253	260	266	272	277	282	286	291	295	298
2	14·0	18·9	22·3	24·7	26·6	28·2	29·5	30·7	31·7	32·6	33·4	34·2	34·8	35·5	36·0	36·5	37·0	37·5	38·0
3	8·26	10·56	12·17	13·34	14·25	15·00	15·65	16·20	16·69	17·13	17·53	17·89	18·23	18·54	18·83	19·09	19·33	19·56	19·79
4	6·51	8·08	9·17	9·97	10·58	11·10	11·55	11·93	12·26	12·56	12·84	13·09	13·32	13·53	13·73	13·92	14·09	14·25	14·40
5	5·62	6·83	7·65	8·26	8·73	9·12	9·46	9·76	10·02	10·25	10·46	10·65	10·83	10·99	11·14	11·29	11·42	11·54	11·66
6	5·24	6·32	7·03	7·56	7·97	8·31	8·61	8·87	9·10	9·30	9·49	9·65	9·81	9·95	10·08	10·21	10·32	10·43	10·54
7	4·94	5·89	6·52	6·98	7·35	7·65	7·91	8·14	8·34	8·52	8·68	8·82	8·96	9·08	9·20	9·31	9·42	9·52	9·61
8	4·74	5·63	6·20	6·63	6·96	7·24	7·47	7·68	7·86	8·03	8·18	8·31	8·44	8·55	8·66	8·76	8·86	8·95	9·03
9	4·60	5·42	5·96	6·35	6·66	6·91	7·13	7·33	7·50	7·65	7·79	7·91	8·02	8·13	8·23	8·33	8·41	8·50	8·58
10	4·48	5·26	5·77	6·14	6·43	6·67	6·88	7·06	7·22	7·36	7·49	7·60	7·71	7·81	7·91	7·99	8·08	8·15	8·23
11	4·39	5·14	5·62	5·98	6·25	6·47	6·67	6·84	6·99	7·13	7·25	7·36	7·46	7·56	7·65	7·73	7·81	7·88	7·95
12	4·32	5·04	5·50	5·84	6·10	6·32	6·51	6·67	6·81	6·94	7·06	7·17	7·26	7·36	7·44	7·52	7·60	7·67	7·73
13	4·26	4·96	5·40	5·73	5·98	6·19	6·37	6·53	6·67	6·79	6·90	7·01	7·10	7·19	7·27	7·35	7·42	7·49	7·55
14	4·21	4·89	5·32	5·64	5·88	6·08	6·26	6·41	6·54	6·66	6·77	6·87	6·96	7·05	7·13	7·20	7·27	7·34	7·40
15	4·17	4·83	5·25	5·56	5·80	5·99	6·16	6·31	6·44	6·55	6·66	6·76	6·85	6·93	7·00	7·07	7·14	7·20	7·26
16	4·13	4·78	5·19	5·49	5·72	5·91	6·08	6·22	6·35	6·46	6·56	6·66	6·74	6·82	6·90	6·97	7·03	7·09	7·15
17	4·10	4·73	5·14	5·43	5·66	5·85	6·01	6·15	6·27	6·38	6·48	6·57	6·66	6·73	6·81	6·87	6·94	7·00	7·05
18	4·07	4·70	5·09	5·38	5·60	5·79	5·95	6·08	6·20	6·31	6·41	6·50	6·58	6·65	6·73	6·79	6·85	6·91	6·97
19	4·05	4·66	5·05	5·34	5·55	5·73	5·89	6·02	6·14	6·25	6·34	6·43	6·51	6·58	6·65	6·72	6·78	6·84	6·89
20	4·02	4·63	5·02	5·30	5·51	5·69	5·84	5·97	6·09	6·19	6·28	6·37	6·45	6·52	6·59	6·66	6·71	6·77	6·82
24	3·96	4·54	4·91	5·17	5·37	5·54	5·69	5·81	5·92	6·02	6·11	6·19	6·26	6·33	6·39	6·45	6·51	6·57	6·61
30	3·89	4·45	4·80	5·05	5·24	5·40	5·53	5·65	5·76	5·85	5·93	6·01	6·08	6·14	6·20	6·26	6·31	6·36	6·41
40	3·82	4·36	4·70	4·93	5·11	5·26	5·39	5·50	5·60	5·69	5·77	5·84	5·90	5·96	6·02	6·07	6·12	6·17	6·21
60	3·76	4·28	4·60	4·82	4·99	5·13	5·25	5·36	5·45	5·53	5·60	5·67	5·73	5·78	5·83	5·88	5·93	5·98	6·01
120	3·70	4·20	4·50	4·71	4·87	5·00	5·12	5·21	5·30	5·38	5·44	5·50	5·56	5·61	5·66	5·71	5·75	5·79	5·83
∞	3·64	4·12	4·40	4·60	4·76	4·88	4·99	5·08	5·16	5·23	5·29	5·35	5·40	5·45	5·49	5·53	5·57	5·61	5·64

On the distribution of 'Studentized' range

By K. C. S. PILLAI, *University of Travancore, Trivandrum*

1. Introduction

Let $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ denote an ordered sample of size n and $X_1, X_2, \ldots, X_{\nu+1}$ a second independent sample of size $\nu+1$ from a normal population. Then the 'Studentized' range is defined as

$$q = \frac{w}{s},$$

where $w = x_n - x_1$ and $s^2 = \sum_{1}^{\nu+1} (X_i - \bar{X})^2/\nu$. It may be mentioned that q is a particular case of the general class of 'Studentized' functions discussed by H. O. Hartley (1938). The usefulness of q in practical problems has been illustrated by Newman (1939), and by Pearson & Hartley (1943). Newman (1939) obtained by quadrature, using the approximate probability law of w due to Pearson, the 5 and 1 % upper probability levels of q for small values of n and $\nu \geqslant 5$. Pearson & Hartley (1943) revised the table of probability levels given by Newman with the help of the exact tables of the probability integral for w (Hartley, 1942; Pearson & Hartley, 1942); they have calculated the upper and lower percentage points of q for values of n ranging from 2 to 20 and $\nu \geqslant 10$ and also the probability integral of q for the same values of n and ν based on certain results obtained by Hartley (1944). In the present paper a study of the distribution of q has been based on the distribution of semi-range developed as a series (Pillai 1948, 1951) and the percentage points have been calculated for small values of n and ν, to fill the gap in the tables prepared by the former authors. The lower percentage points are given below for values of n ranging from 2 to 8 and $\nu < 10$; the upper percentage points for the same values of n and $\nu < 5$ were also calculated.*

2. The distribution of semi-range

It has been shown (Pillai, 1951) that the distribution of semi-range, $W = \frac{1}{2}(x_n - x_1)$ can be obtained in the form

$$p(W) = kW^{n-2} e^{-\frac{1}{2}(n+4)W^2} F, \tag{1}$$

where

$$k = \frac{n-1}{2^4 n^{\frac{1}{2}} \Gamma(5)} \left(\frac{2}{\pi}\right)^{\frac{1}{2}(n-1)}$$

and

$$F = C_0^{(n)} + C_1^{(n)} W^2 + C_2^{(n)} W^4 + \ldots, \tag{2}$$

the C's being functions of n. The $C^{(n)}$ coefficients for values of n ranging from 3 to 8 have been calculated and given in Table 1.

Table 1. *$C^{(n)}$ coefficients for values of n ranging from* 3 *to* 8

n	$C_0^{(n)}$	$C_1^{(n)}$	$C_2^{(n)}$	$C_3^{(n)}$	$C_4^{(n)}$	$C_5^{(n)}$	$C_6^{(n)}$
3	31104	1728	201·6000	6·1714	0·3397	0·0042	
4	98304	8192	1911·4667	42·8021	14·8634	− 0·6523	
5	240000	24000	7920·0000	311·7460	94·8921	− 4·1893	2·4859
6	497664	55296	23040·0000	1243·4287	398·7097	− 16·7465	13·7155
7	921984	109760	54618·6667	3565·9259	1263·4199	− 42·1564	48·9035
8	1572864	196608	113049·6000	8374·0445	3293·8665	− 73·9556	136·3427

* [Mr Pillai's paper was submitted for publication when the more extensive programme of computation described on pp. 192–3 of this issue of the journal was already in hand. His new upper percentage points are not therefore printed separately as they are included in the main body of the tables on pp. 192–3. Ed.]

3. The distribution of the 'Studentized' range

Starting with the distribution law (1) of W and the distribution of s given by

$$p(s) = \frac{\nu^{\frac{1}{2}\nu}}{2^{\frac{1}{2}(\nu-2)}} \frac{1}{\Gamma(\frac{1}{2}\nu)} s^{\nu-1} e^{-\frac{1}{2}\nu s^2}, \tag{3}$$

putting $q = 2W/s$ in $p(W).p(s)$, and integrating out s in the interval 0 to ∞, the distribution of q may be easily seen to be of the form

$$p(q) = \frac{k\nu^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} \sum_{i=0}^{\infty} C_i^{(n)} \frac{q^{n+2i-2}}{2^{\frac{1}{2}(n+2i-1)}} \frac{\Gamma\frac{1}{2}(n+2i+\nu-1)}{\{\frac{1}{12}(n+4)\,q^2+\nu\}^{\frac{1}{2}(n+2i+\nu-1)}}. \tag{4}$$

As the terms in (4) reduce to Beta-functions, the percentage points can be evaluated with the help of *Tables of the Incomplete Beta-function*.

4. Percentage points of q

Using (4), upper and lower percentage points of q were computed for small values of n and ν, for the 5 and 1 % significance levels; the lower percentage points are presented in Tables 2 and 3.

Table 2. *Lower 5 % points of the 'Studentized' range*

ν \ n	2	3	4	5	6	7	8
1	0·11	0·44	0·70	0·88	1·02	1·14	1·23
2	0·10	0·44	0·72	0·93	1·09	1·22	1·33
3	0·10	0·44	0·73	0·95	1·13	1·27	1·39
4	0·09	0·43	0·73	0·97	1·15	1·30	1·43
5	0·09	0·43	0·74	0·98	1·17	1·32	1·45
6	0·09	0·43	0·74	0·99	1·18	1·34	1·47
7	0·09	0·43	0·75	0·99	1·19	1·35	1·48
8	0·09	0·43	0·75	1·00	1·19	1·36	1·50
9	0·09	0·43	0·75	1·00	1·20	1·37	1·51

Table 3. *Lower 1 % points of the 'Studentized' range*

ν \ n	2	3	4	5	6	7	8
1	0·02	0·19	0·38	0·53	0·65	0·75	0·83
2	0·02	0·19	0·40	0·58	0·72	0·84	0·93
3	0·02	0·19	0·41	0·60	0·75	0·88	0·99
4	0·02	0·19	0·41	0·61	0·77	0·91	1·03
5	0·02	0·19	0·42	0·62	0·79	0·93	1·05
6	0·02	0·19	0·42	0·63	0·80	0·95	1·07
7	0·02	0·19	0·42	0·63	0·81	0·96	1·09
8	0·02	0·19	0·42	0·64	0·81	0·96	1·10
9	0·02	0·19	0·42	0·64	0·81	0·97	1·11

REFERENCES

Hartley, H. O. (1938). *J.R. Statist. Soc. Suppl.* **5**, 80.
Hartley, H. O. (1942). *Biometrika*, **32**, 334.
Hartley, H. O. (1944). *Biometrika*, **33**, 173.
Newman, D. (1939). *Biometrika*, **31**, 20.
Pearson, E. S. & Hartley, H. O. (1942). *Biometrika*, **32**, 301.
Pearson, E. S. & Hartley, H. O. (1943). *Biometrika*, **33**, 89.
Pillai, K. C. S. (1948). *Sankhyā*, **8**, 375.
Pillai, K. C. S. (1951). *Sankhyā*, **11**, 23.

Note on a certain family of discrete distributions

By J. S. MARITZ

National Institute for Personnel Research, South African Council for Scientific and Industrial Research

1. We wish to distinguish between 'addition' and 'superimposition' of probability distributions and will define the terms as follows:

If we have a number of probability distributions $p_1(x_1)$, $p_2(x_2)$, ..., then $p(x) = p_1(x_1) + p_2(x_2) + \dots$ will be the 'superimposition' of these distributions, and $p(x) = p(x_1 + x_2 + \dots)$ will be their 'addition'.

2. We will consider the distribution obtained by the addition of two correlated Poisson variables, since the 'model' of §3 is a direct extension of this model to k variates.

Consider two events A and B of which only the following (mutually exclusive) combinations may happen during an interval of time δt:

$$A\bar{B}, \quad B\bar{A}, \quad AB, \quad \bar{A}\bar{B},$$

and let the probabilities associated with these possible (joint) events be

$$a\,\delta t, \quad b\,\delta t, \quad c\,\delta t, \quad 1-(a+b+c)\,\delta t.$$

Let the random variables associated with A and B be x and y, and let the joint probability of A occurring x times and B y times during the interval 0 to t be $p(x, y, t)$. Then

$$p(x, y, t+\delta t) = p(x-1, y, t)\,a\,\delta t + p(x, y-1, t)\,b\,\delta t + p(x-1, y-1, t)\,c\,\delta t + p(x, y, t)\,(1-(a+b+c)\,\delta t). \quad (1)$$

Defining

$$\phi_{xy}(\theta_1, \theta_2, t) = \sum_{x,y=0}^{\infty} \theta_1^x \theta_2^y p(x, y, t), \quad (2)$$

we have

$$\phi_{xy}(\theta_1, \theta_2, t+\delta t) = \phi_{xy}(\theta_1, \theta_2, t)\,\theta_1 a\,\delta t + \phi_{xy}(\theta_1, \theta_2, t)\,\theta_2 b\,\delta t + \phi_{xy}(\theta_1, \theta_2, t)\,\theta_1\theta_2 c\,\delta t + \phi_{xy}(\theta_1, \theta_2, t)\,(1-(a+b+c)\,\delta t), \quad (3)$$

so that

$$\frac{1}{\phi}\frac{\delta\phi}{\delta t} = a\theta_1 + b\theta_2 + c\theta_1\theta_2 - (a+b+c). \quad (4)$$

Whence, letting $\delta t \to 0$ and since at $t = 0$, $p(0, 0, t) = 1$ and $p(x, y, t) = 0$ for $x, y \neq 0$,

$$\phi_{xy}(\theta_1, \theta_2, t) = \exp\{(a\theta_1 + b\theta_2 + c\theta_1\theta_2 - (a+b+c))\,t\}. \quad (5)$$

Now clearly the total probability of A in δt is

$$(a+c)\,\delta t,$$

so that x is a Poisson variable with parameter $(a+c)\,t$. This can also be seen by putting $\theta_2 = 1$ in (5). Similarly y is a Poisson variable with parameter $(b+c)\,t$, and $\operatorname{cov}(x, y) = ct$. Putting $\theta_1 = \theta_2 = \theta$ in (5) gives the 'probability generating function' (p.g.f.) of $z = x + y$,

$$\phi_z(\theta, t) = \exp\{(c\theta^2 + (a+b)\,\theta - (a+b+c))\,t\}. \quad (6)$$

A description of the bivariate Poisson distribution leading to (5) is given in Aitken (1944).

3. As an extension of the model of §2, consider n events $e_1, e_2, \dots, e_n$ of which any combination of the type $e_1, \bar{e}_2, \bar{e}_3, e_4, \dots, \bar{e}_n$ (for example), which contains every e_r only once, may occur during the interval of time δt. (These $\sum_{r=0}^{n} \binom{n}{r}$ possible combinations are assumed mutually exclusive.) Let the probability associated with the occurrence of such a combination, denoted by $\bar{e}$, say, be $a_{\bar{e}}\,\delta t$, and let the probability of any combination which contains any e_r more than once be $= O(\delta t^2)$.

Let the random variables associated with $e_1 \dots e_n$ be $x_1 \dots x_n$ and denote $(x_1 \dots x_n)$ by X. Clearly, the total probability of occurrence of e_i, say, during δt is $(\sum_{\bar{e}_i} a_{\bar{e}})\,\delta t + O(\delta t^2)$, where $\sum_{\bar{e}_i}$ denotes summation over all possible $\bar{e}$ containing e_i, so that x_i is a Poisson variable.

We now wish to determine the joint distribution of the variables $x_1 \dots x_n$. Writing the joint probability of e_1 occurring x_1 times, e_2 occurring x_2 times, etc., during the time interval 0 to t as $p(x_1 \dots x_n, t)$ we define the joint p.g.f. of $x_1 \dots x_n$ by

$$\phi_{x_1 \dots x_n}(\theta_1 \dots \theta_n, t) = \phi_x(\theta_1 \dots \theta_n, t) = \Sigma\theta_1^{x_1} \dots \theta_n^{x_n} p(x_1 \dots x_n, t).$$

Let $(\bar{x})$ be a vector containing 1's corresponding to the e_r in $\bar{e}$ and 0's corresponding to the $\bar{e}_r$ in $\bar{e}$. The following scheme will make this clear:

$$\left.\begin{aligned}(\bar{e}) &= (e_1, \bar{e}_2, e_3, e_4 \dots \bar{e}_n),\\ (\bar{x}) &= (1, 0, 1, 1, \dots, 0),\end{aligned}\right\} \tag{7}$$

so that $$(X)-(\bar{x}) = (\bar{X}) = (x_1-1, x_2, x_3-1, x_4-1, \dots, x_n).$$

Let $\theta_{\bar{e}}$ denote the product of all the θ_i corresponding to the e_i in $\bar{e}$. We then have

$$p(X, t+\delta t) = \sum_{\bar{e}} p(\bar{X}, t)\, a_{\bar{e}}\delta t + p(X, t)\,(1-\sum_{\bar{e}} a_{\bar{e}}\delta t) + O(\delta t^2), \tag{8}$$

where $\sum_{\bar{e}}$ denotes summation over all combinations $\bar{e}$, whence

$$\phi(\theta_1 \dots \theta_n, t+\delta t) - \phi(\theta_1 \dots \theta_n, t) = \sum_{\bar{e}} \phi(\theta_1 \dots \theta_n, t)\,\theta_{\bar{e}} a_{\bar{e}}\delta t - \sum_{\bar{e}} \phi(\theta_1 \dots \theta_n, t)\, a_{\bar{e}}\delta t + O(\delta t^2). \tag{9}$$

Letting $\delta t \to 0$ we have $$\frac{1}{\phi(\theta_1 \dots \theta_n, t)}\frac{d\phi(\theta_1 \dots \theta_n, t)}{dt} = \sum_{\bar{e}} (\theta_{\bar{e}} a_{\bar{e}} - a_{\bar{e}}), \tag{10}$$

and since at $t = 0, p(0 \dots 0, t) = 1$ and $p(x_1 \dots x_n, t) = 0$ for any $x_r \neq 0$,

$$\phi(\theta_1 \dots \theta_n, t) = \exp\{t \sum_{\bar{e}} (\theta_{\bar{e}} a_{\bar{e}} - a_{\bar{e}})\}. \tag{11}$$

(11) defines the joint p.g.f. of n variables each of which has a Poisson distribution. As can be seen quite easily by putting $n = 2$ these n variables are in general not independent.

By putting $\theta_1 = \theta_2 = \dots = \theta_n = \theta$ we obtain the p.g.f. of the variable $z = x_1 + x_2 + \dots + x_n$. We thus obtain a distribution which is the *addition* of a set of *correlated* Poisson variables as contrasted with the super-imposition of independent Poisson distributions which is encountered quite commonly, especially in the literature concerning 'Accident Proneness'.

From (11), the p.g.f. $\phi_z(\theta, t)$ of z is obtained as

$$\left.\begin{aligned}\phi_z(\theta, t) &= \exp\{t\Sigma(a_r\theta^r - a_r)\}\\ &= \exp\{\Sigma(\alpha_r\theta^r - \alpha_r)\}\\ &= \phi(\theta).\end{aligned}\right\} \tag{12}$$

From (12) we have $$p(z) = \exp\Big[-\sum_{r=1}^{n}\alpha_r\Big]\sum_{s=1}^{*z}\frac{1}{s!}\sum_{\pi_1 \dots \pi_s=1}^{n}\alpha_{\pi_1}\dots\alpha_{\pi_s}, \tag{13}$$

where $\pi_1 + \dots + \pi_s = z$ and each $\pi_i \geqslant 1$. More explicitly

$$\left.\begin{aligned}p(0) &= e^{-\Sigma\alpha_r},\\ p(1) &= e^{-\Sigma\alpha_r}\alpha_1,\\ p(2) &= e^{-\Sigma\alpha_r}\Big(\frac{\alpha_1^2}{2!} + \alpha_2\Big),\\ p(3) &= e^{-\Sigma\alpha_r}\Big(\frac{\alpha_1^3}{3!} + \alpha_1\alpha_2 + \alpha_3\Big),\\ p(4) &= e^{-\Sigma\alpha_r}\Big(\frac{\alpha_1^4}{4!} + \frac{\alpha_2^2}{2} + \alpha_1\alpha_3 + \frac{\alpha_1^2\alpha_2}{2} + \alpha_4\Big),\end{aligned}\right\} \tag{14}$$

etc.

It should be noted that if $n \to \infty$ a certain class of distributions is obtained, and it may be of interest to try to determine which distributions belong to this class. Certainly those of which the $\log\phi(\theta)$ can be expanded in an infinite Maclaurin series with radius of convergence $R > 1$, for

$$\psi(\theta) = \log\phi(\theta) = 0 \quad \text{at} \quad \theta = 1,$$

so that $$\psi(0) = -\sum_{r=1}^{\infty}\frac{\psi_{(0)}^{(r)}}{r!},$$

and $$\phi(\theta) = \exp\left\{\Sigma\left(\frac{\psi_{(0)}^{(r)}}{r!}\theta^r - \frac{\psi_{(0)}^{(r)}}{r!}\right)\right\}. \tag{15}$$

which is of the same form as (12).

The negative binomial distribution belongs to this class as, for this p.d.,

$$\phi(\theta) = \left(\frac{c}{c+1}\right)^p \left(1 - \frac{\theta}{c+1}\right)^{-p},$$

where $c, p > 0$, and

$$\psi(\theta) = p \log \frac{c}{c+1} - p \log \left(1 - \frac{\theta}{c+1}\right),$$

which can be expanded in a Maclaurin series for $\theta < c+1$.

4. The factorial cumulant generating function of the distribution (13) is given as

$$g(\tau) = \log \phi(1+\tau) \tag{16}$$

$$= \tau \sum_{r=1}^{n} r\alpha_r + \frac{\tau^2}{2!} \sum_{r=2}^{n} r(r-1)\alpha_r + \ldots + \tau^n \alpha_n, \tag{17}$$

so that

$$\kappa_{[s]} = \sum_{r=s}^{n} r(r-1)\ldots(r-s+1)\alpha_r \tag{18}$$

and if $(\kappa_{[s]}) = M(\alpha_s)$, where M is a matrix, then M is diagonal, and so is M^{-1}. This gives some simple relations which may be used to estimate the α_r, e.g. if $n = 3$,

$$\left.\begin{aligned} \alpha_1 &= \kappa_{[1]} - \kappa_{[2]} + \tfrac{1}{2}\kappa_{[3]}, \\ \alpha_2 &= \tfrac{1}{2}\kappa_{[2]} - \tfrac{1}{2}\kappa_{[3]}, \\ \alpha_3 &= \frac{1}{3!}\kappa_{[3]}. \end{aligned}\right\} \tag{19}$$

5. *Practical considerations.* (*a*) A possible practical application of the distribution (13) is to the so-called '*c* chart' (see Grant, 1946, p. 295). Current practice is to use control limits based on the Poisson distribution. This presupposes that the defects of the parts making up the unit occur independently. This need not always be so, in which case (13) should be applicable—with the appropriate number of parameters. Unfortunately we have no such data to which to apply the theory.

(*b*) It is felt that a distribution of type (13) may be useful in certain problems of absenteeism in which the absences of all individuals are not independent events. Again we have no data in support of this suggestion, but it is hoped that some will be available soon.

The author wishes to express his gratitude to the South African Council for Scientific and Industrial Research for permission to publish this paper.

REFERENCES

AITKEN, A. C. (1944). *Statistical Mathematics.* University Mathematical Texts, 2. Edinburgh and London: Oliver and Boyd.

GRANT, E. L. (1946). *Statistical Quality Control.* New York and London: McGraw-Hill Book Co. Inc.

Some properties of runs in smoothed random series

BY ALISON M. GRANT, *Department of Meteorology, University of Melbourne*

The method of smoothing a series by means of a moving average with equal weights is often used to reduce short-period fluctuations in series of experimental data. However, in order to draw valid conclusions concerning the data from the series obtained by this process, a knowledge of the effects of such a technique on random series is required. The present paper gives the solution of this problem for some properties of runs, a run being defined as a series of values showing a continuous increase or decrease. In particular, the frequency distribution of runs of different lengths is obtained as a function of the extent of the moving average. Amplitudes, i.e. differences between highest and lowest values, of runs of different lengths for both random series and derived series of averages, are also considered.

Distribution of runs of different lengths

The distribution of runs in a random series has been studied by Kermack & McKendrick (1937); the derivation, in rather a different form, is given here both in order to illustrate the method and because it forms part of the study of runs in the smoothed series.

If the values of a continuous variate x are chosen at random from a population with probability distribution function $f(x)$, the function y, defined by the relation

$$y = \int_{-\infty}^{x} f(x)\,dx,$$

will be distributed in a rectangular distribution in the range (0, 1). There is a one-to-one correspondence between values of x and values of y, so that the inequality $x_1 < x_2$ implies the inequality $y_1 < y_2$, etc.

The conditions which must be satisfied for a particular point, say x_0, to be the start of an upward run of length n, consisting of n consecutive upward steps preceded and succeeded by downward steps, are given by

$$x_{-1} > x_0 < x_1 < x_2 < \ldots < x_n > x_{n+1},$$

or alternatively by

$$y_{-1} > y_0 < y_1 < y_2 < \ldots < y_n > y_{n+1}.$$

The probability that these conditions will be satisfied is given by

$$\int_0^1 \int_{y_{n+1}}^1 \int_0^{y_n} \ldots \int_0^{y_2} \int_0^{y_1} \int_{y_0}^1 dy_{-1}\,dy_0\,dy_1 \ldots dy_{n-1}\,dy_n\,dy_{n+1}, \tag{1}$$

which is equal to

$$\frac{n^2+3n+1}{(n+3)!}.$$

The probability of occurrence of a run of length n (up or down) is therefore given by

$$P_n = \frac{2(n^2+3n+1)}{(n+3)!}.$$

The result $P_0 = \frac{1}{3}$ indicates the expected frequency of points which do not start a run, i.e. in which $x_{-1} < x_0 < x_1$ or $x_{-1} > x_0 > x_1$. As the starting point of a run of length n is followed by $n-1$ such points, we have

$$\sum_{n=1}^{\infty} (n-1)P_n = P_0,$$

a result which may be used to determine the expected length of runs

$$\mathscr{E}(n) = \sum_{n=1}^{\infty} nP_n \Big/ \sum_{n=1}^{\infty} P_n = \frac{1}{1-P_0} = 1{\cdot}5.$$

Turning now to the results to be expected from a moving average of extent a, the typical term z_r being defined by

$$az_r = x_r + x_{r+1} + \ldots + x_{r+a-1},$$

where, as before, the values of x are chosen in a random manner, we find that the inequality $z_{r+1} > z_r$ requires $x_{r+a} > x_r$ and hence $y_{r+a} > y_r$. The conditions which must be satisfied for a run of a certain length to occur now concern points which are separated by an interval a. It is convenient therefore to use $ma+s$ for the length of the run, where $m = 0, 1, 2, \ldots$; $s = 0, 1, 2, \ldots, a-1$.

As the form of the conditions which must be satisfied by y for a run of length n (in z) depends on the value of s, it is simpler to consider the probability of occurrence of a run of length greater than or equal to $ma+s$. The inequalities

$$z_{-1} > z_0 < z_1 < \ldots < z_{ma+s}$$

imply

$$\left.\begin{array}{cccc}
 & y_s & < y_{a+s} & < \ldots < y_{ma+s} \\
 & y_{s+1} & < y_{a+s+1} & < \ldots < y_{ma+s+1} \\
 & \vdots & & \\
 & y_{a-2} & < y_{2a-2} & < \ldots < y_{ma+a-2} \\
y_{-1} > & y_{a-1} & < y_{2a-1} & < \ldots < y_{ma+a-1} \\
y_0 < & y_a & < y_{2a} & < \ldots < y_{ma+a} \\
y_1 < & y_{a+1} & < y_{2a+1} & < \ldots < y_{ma+a+1} \\
 & \vdots & & \\
y_{s-1} < & y_{a+s-1} & < y_{2a+s-1} & < \ldots < y_{ma+a+s-1}.
\end{array}\right\} \tag{2}$$

Each of the first $a-s-1$ lines leads, by a process of successive integration, to the probability $\frac{1}{(m+1)!}$; the next line gives $\frac{1}{(m+1)!} - \frac{1}{(m+2)!}$; while each of the remaining s lines leads to $\frac{1}{(m+2)!}$. The probability that all of the conditions will be satisfied is therefore

$$\left\{\frac{1}{(m+1)!}\right\}^{a-s-1} \left\{\frac{1}{(m+1)!} - \frac{1}{(m+2)!}\right\} \left\{\frac{1}{(m+2)!}\right\}^{s},$$

which, when doubled to allow for both upward and downward runs, gives

$$P(n \geqslant ma+s) = \frac{2(m+1)}{\{(m+1)!\}^a (m+2)^{s+1}}. \tag{3}$$

The probability of occurrence of a run of length $ma+s$ is obtained as the difference between $P(n \geqslant ma+s)$ and $P(n \geqslant ma+s+1)$.

The result includes that for the random series, which can be considered as a special case with the extent of the moving average equal to unity. For $a>1$, $P_0 = \frac{1}{2}$, showing that the number of peaks and troughs is reduced from two-thirds to one-half of the total number of points by the averaging process. The mean length of runs is now

$$\mathscr{E}(n) = \frac{1}{1-P_0} = 2,$$

which is independent of the extent of the moving average (provided $a>1$).

Relative frequencies of runs of different lengths are shown in Table 1 for the values $a = 1$, 3, and 5. If 'runs' of length 0 are not considered, the frequencies for $a = 1$ have to be multiplied by 1·5, the others by 2; but the numbers as they stand give the expected frequencies in terms of the total length of the series. The values for $a \geqslant 9$ are derived from the fact that the relative frequencies up to $n = a-2$ are given by $\frac{1}{2^{n+1}}$. The increase in the expected frequencies of longer runs as a increases is clearly shown, the value $n = a$ being specially preferred.

Table 1. *Relative frequencies of runs of various lengths* (× 100)

n	$a = 1$	$a = 3$	$a = 5$	$a \geqslant 9$
0	33·33	50·00	50·00	50·00
1	41·67	25·00	25·00	25·00
2	18·33	8·33	12·50	12·50
3	5·28	11·11	6·25	6·25
4	1·15	3·71	2·09	3·13
5	0·21	1·16	2·77	1·56
6	0·03	0·52	0·93	0·78
7	0·00	0·13	0·31	0·39
$\geqslant 8$	0·00	0·04	0·15	0·39

Mean amplitudes of runs

By performing the integration (1) with a fixed value of y_n it is possible to obtain the distribution of y_n for a given value of n. Thus (1) becomes

$$\int_0^{y_n}\int_0^{y_n}\int_0^{y_{n-1}} \cdots \int_0^{y_1}\int_0^{y_1}\int_{y_0}^{1} dy_{-1}dy_0dy_1 \ldots dy_{n-2}dy_{n-1}dy_{n+1},$$

which is equal to

$$\frac{y_n^{n+1}}{n!} - \frac{y_n^{n+2}}{(n+1)!}.$$

Since y, apart from the conditions imposed by the runs, is distributed in a rectangular distribution in the range (0, 1), the conditional distribution of y_n, the end-value in an upward run of length n, is given by

$$f(y_n)\,dy_n = \frac{2}{P_n}\left\{\frac{y_n^{n+1}}{n!} - \frac{y_n^{n+2}}{(n+1)!}\right\}dy_n.$$

If the distribution of x is of such a form that y can be expressed as a function of x, not involving integrals, this leads directly to the distribution of x_n. In any case the expected values of x_n and x_0, and hence of the amplitude, can be found numerically if not in functional form. Thus

$$\mathscr{E}(x_n) = \frac{2}{P_n}\int_0^1 x\left\{\frac{y^{n+1}}{n!} - \frac{y^{n+2}}{(n+1)!}\right\}dy,$$

where x is expressed in terms of y.

The expected value of each of the a values of x which go to make up z_n, the last value in a run in the derived series, is found in a similar manner from the conditions (2). The additional condition for the

termination of the run, $y_{ma+s} > y_{ma+a+s}$, appears at the end of the first line of (2), but it must be borne in mind that if $s = a-1$, this is of the type

$$y_{-1} > y_{a-1} < y_{2a-1} < \dots < y_{ma+a-1} > y_{ma+2a-1},$$

which leads to the same results as the random case with m replacing n. The distributions of the corresponding y values (apart from constant multipliers) are given in Table 2, together with the number of terms which follow the law in question.

Table 2

(i) $s < a-1$

No.	$f(y)$
1	$(m+1)\,y^m - y^{m+1}$
$s+1$	y^{m+1}
$a-s-2$	y^m

(ii) $s = a-1$

No.	$f(y)$
1	$(m+1)\,y^{m+1} - y^{m+2}$
$a-1$	y^{m+1}

For each type,

$$\mathscr{E}(x) = \int_0^1 x f(y)\,dy \Big/ \int_0^1 f(y)\,dy,$$

and $\mathscr{E}(z_n)$ is equal to the average of the a values of $\mathscr{E}(x)$. Substitution of the value of x corresponding to $1-y$ leads in a similar manner to $\mathscr{E}(z_0)$, and hence to the expected value of the amplitude, $\mathscr{E}(z_n - z_0)$.

The simplest example is where x is distributed in a rectangular distribution in the range $(0, 1)$. Here $x = y$ and the various integrals can be directly evaluated. The variation of mean amplitude with length of run is shown by the heavy lines in Fig. 1 for the cases $a = 1, 3$ and 5. The amplitudes are expressed in terms of σ_x, the standard deviation of x, so that they can be compared with corresponding results obtained from other parent distributions. The factor a has been introduced in order to give a clearer picture of the pattern of variation.

There is a marked increase in the importance of the longer runs in the smoothed series as compared with the random series ($a = 1$). Thus the averaging process not only increases the expected frequency of long runs, but also increases their amplitudes relative to that of short runs.

The mean amplitudes of runs were also calculated for x distributed in a standard normal distribution. Their dependence on the extent of the average and on the length of run is indicated by the fine lines in Fig. 1. The values for the smoothed series ($a = 3$ and $a = 5$) are practically identical with those found from the rectangular distribution. For the random series ($a = 1$) there is a slight increase in the amplitudes of the longer runs over those obtained from the rectangular distribution, due to the removal of the absolute limitation on the values of x.

This comparison suggests that the results obtained from the rectangular distribution give a fairly accurate picture of the behaviour of runs in series derived from other parent distributions. It is therefore of interest to carry the derivation a stage further in order to obtain the distributions of the amplitudes as well as the mean amplitude for given n.

Distribution of amplitudes

The joint distribution of y_0 and y_n (for the random series) is obtained by integrating (1) with fixed values of y_0 and y_n. The resulting expression

$$\int_0^{y_n} \int_{y_0}^{y_n} \int_{y_0}^{y_{n-1}} \dots \int_{y_0}^{y_3} \int_{y_0}^{y_2} \int_{y_0}^{1} dy_{-1}\,dy_1\,dy_2 \dots dy_{n-2}\,dy_{n-1}\,dy_{n+1}$$

is equal to

$$y_n(1-y_0)\frac{(y_n - y_0)^{n-1}}{(n-1)!}.$$

As the unconditional distribution of y is rectangular, the joint conditional distribution of y_0 and y_n is given by

$$f(y_0, y_n)\,dy_0 dy_n = \frac{2}{P_n} y_n(1-y_0)\frac{(y_n-y_0)^{n-1}}{(n-1)!}\,dy_0 dy_n.$$

Hence

$$f(y_0, A_n)\,dy_0 dA_n = \frac{2}{P_n}(y_0+A_n)(1-y_0)\frac{A_n^{n-1}}{(n-1)!}dy_0 dA_n,$$

where $A_n = y_n - y_0$. The marginal distribution of A_n is given by

$$f(A_n)\,dA_n = \frac{1}{P_n}\frac{A_n^{n-1}}{3(n-1)!}\{1+3A_n-3A_n^2-A_n^3\}\,dA_n.$$

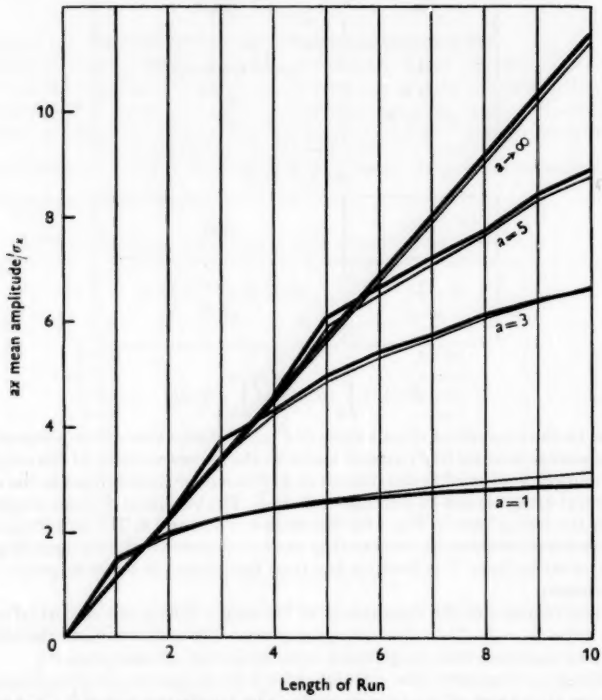

Fig. 1. Variation of mean amplitude with length of run for a moving average of extent a.

——— Rectangular distribution of x; ——— Normal distribution of x.

If x is distributed in a rectangular distribution in the range (0, 1), A_n is equal to the amplitude of a run of length n and the above function represents its probability distribution.

In the case of the moving average the distribution of the difference between each pair of y values, one of which corresponds to one of the x values forming part of z_0, the other to one forming part of z_n, can be found in the same way. The different types of distribution of A which arise in this manner are given in Table 3, together with the number of times they occur and also their means and variances.

For particular values of a and n it is possible to derive the distribution of the amplitude, $z_n - z_0$, which is equal to the mean of the a values of A chosen from the appropriate distributions. Thus for a moving average of extent 2, and runs of length 2, so that $m = 1$ and $s = 0$, the joint distribution of the two values of A is given by $f(A_1, A_2)\,dA_1 dA_2 = \frac{9}{4}(1-A_1^2)(1-A_2^2)\,dA_1 dA_2$.

It is readily shown that the distribution function of $B = \frac{1}{2}(A_1 + A_2)$ takes the form

$$f(B)\,dB \begin{cases} = \frac{3}{5}B(8B^4 - 40B^2 + 15)\,dB, & B < \frac{1}{2}; \\ = \frac{2}{5}4(1-B)^3(B^2+3B+1)\,dB, & B > \frac{1}{2}. \end{cases}$$

It does not seem likely that the amplitude distribution can be given simply in a general form. However, as the amplitude is the mean of a independent variates, it will be approximately normally distributed with mean and variance found from the a individual values, i.e.

$$\mathscr{E}(z_n - z_0) = \frac{\Sigma \mathscr{E}(A)}{a}, \quad V(z_n - z_0) = \frac{\Sigma V(A)}{a^2}.$$

The numerical values of $\mathscr{E}(A)$ and $V(A)$ are given in Table 4 for the first few values of m. They can be used to determine the mean and the standard deviation of the amplitude of runs of any given length in

Table 3

(i) $s < a-1$

No.	$f(A)$	$\mathscr{E}(A)$	$V(A)$
2	$A^{m-1} - A^{m+1}$	$\frac{m(m+2)}{(m+1)(m+3)}$	$\frac{m(2m^2+8m+9)}{(m+1)^2(m+3)^2(m+4)}$
s	$A^m - A^{m+1}$	$\frac{m+1}{m+3}$	$\frac{2(m+1)}{(m+3)^2(m+4)}$
$a-s-2$	$A^{m-1} - A^m$	$\frac{m}{m+2}$	$\frac{2m}{(m+2)^2(m+3)}$

(ii) $s = a-1$

No.	$f(A)$	$\mathscr{E}(A)$	$V(A)$
1	$A^{m-1} + 3A^m - 3A^{m+1} - A^{m+2}$	$\frac{m(m+5m+5)}{(m+4)(m^2+3m+1)}$	$\frac{2m(m^4+10m^3+34m^2+45m+22)}{(m+4)^2(m+5)(m^2+3m+1)^2}$
$a-1$	$A^m - A^{m+1}$	$\frac{m+1}{m+3}$	$\frac{2(m+1)}{(m+3)^2(m+4)}$

Table 4. *Values of $\mathscr{E}(A)$ and $V(A)$ (in brackets) for runs of length $ma+s$. (Unit = standard deviation of original series)*

(i) $s < a-1$

No.	$m=0$	$m=1$	$m=2$	$m=3$
2	0	1·30	1·85	2·16
	(0)	(0·71)	(0·59)	(0·45)
s	1·15	1·73	2·08	2·31
	(0·67)	(0·60)	(0·48)	(0·38)
$a-s-2$	0	1·15	1·73	2·08
	(0)	(0·67)	(0·60)	(0·48)

(ii) $s = a-1$

No.	$m=0$	$m=1$	$m=2$	$m=3$
1	0	1·52	1·99	2·27
	(0)	(0·72)	(0·54)	(0·41)
$a-1$	1·15	1·73	2·08	2·31
	(0·67)	(0·60)	(0·48)	(0·38)

a smoothed random series. Thus for a moving average of extent 4 and runs of length 5, so that $m = 1$, $s = 1$, we have

Mean amplitude $= \frac{1}{4}(2 \times 1{\cdot}30 + 1{\cdot}73 + 1{\cdot}15) = 1{\cdot}37$,
Variance of amplitude $= \frac{1}{16}(2 \times 0{\cdot}71 + 0{\cdot}60 + 0{\cdot}67) = 0{\cdot}168$,
Standard deviation $= 0{\cdot}41$.

Both the mean and the standard deviation so obtained are expressed in units equal to the standard deviation of the original random series. For the random series $a = 1$, $s = 0$, so that the amplitude distributions, means, and variances are those appropriate to the single term in (ii).

Discussion

It has been shown that the moving average process applied to a random series gives rise to a series which is far from random in character. In particular, runs of short length are suppressed both in number and in amplitude. Also runs of length equal to the extent of the moving average are increased in number. This is evidence of the oscillatory nature of the moving average which has been studied previously by means of serial correlation coefficients, for example by Spencer-Smith (1947). The dangers of applying the technique when seeking information regarding oscillations in a series must again be emphasized, as the appearance of the data in this form is likely to be misleading.

The results given here may be used as the basis for tests of randomness in a series in which each observation depends on average conditions over a period of time which is longer than the interval between observations. In such a case the series would be similar to a moving average and should be compared with a smoothed random series in order to study its basic properties. The distributions of frequencies and amplitudes of runs could be used to decide whether any fluctuations observed can be explained in this way. In this connexion it should be noted that the amplitudes of consecutive runs are not independent of one another so that the variance of mean amplitude, as calculated, is not strictly true. However, on the hypothesis that the series in question is a moving average based on a random series, one very abnormal value in the latter will affect only a limited number of terms in the former, and hence will affect at most a few runs. The tests suggested would therefore be approximately correct provided the length of the observed series is long compared with the extent of the moving average suspected.

In conclusion, I wish to thank Associate-Professor M. H. Belz and Mr U. Radok for their helpful suggestions during the preparation of this paper.

REFERENCES

Kermack, W. O. & McKendrick, A. G. (1937). *Proc. Roy. Soc. Edinb.* **57**, 228.
Spencer-Smith, J. L. (1947). *J.R. Statist. Soc. Suppl.* **9**, 104.

An approximation to the symmetrical incomplete beta function

By J. H. CADWELL.

1. Summary

The following approximation is developed:

$$I_x(p,p) \simeq \tfrac{1}{2} + L(p)\left\{\Phi(\zeta) + \frac{2(p-1)(2p+1)}{(4p-1)^4}\Psi(\zeta)\right\},$$

$$x = \tfrac{1}{2} + \left(\frac{\pi}{3}\right)^{\frac{1}{2}}\Phi\left\{\left(\frac{3}{4p-1}\right)^{\frac{1}{2}}\zeta\right\}.$$

The functions Φ and Ψ are defined as

$$\Phi(\zeta) = \int_0^\zeta \frac{e^{-\frac{1}{2}t^2}}{\sqrt{(2\pi)}}\,dt, \quad \Psi(\zeta) = \int_0^\zeta \left(1 - \frac{t^6}{15}\right)\frac{e^{-\frac{1}{2}t^2}}{\sqrt{(2\pi)}}\,dt.$$

The latter function is only required to two-place accuracy and a short table is given. The function $L(p)$ is also tabled; for $p > 3$ it can be replaced by unity. This approximation is of use for $p = 2$, where it gives a maximum error of 0·00011; for $p > 5$ it will give five-figure accuracy.

The function $I_x(p,p)$ is needed in evaluating the area under Pearson curves of types II and VII. As $2p$ will in general not be an integer, use of the incomplete beta function tables raises difficulties in interpolation. Thus for four-place accuracy, third differences in p are needed for small p, while for large p second differences in x must be used.

The above method allows the use of four- or five-figure tables of the normal integral with mean differences. It will be found that the saving in time on the interpolation involved is more than enough to compensate for the auxiliary computations needed. The following relations indicate the range of symmetrical Pearson curves covered:

$$p>2, \quad 2\tfrac{1}{7} < \beta_2 < \infty;$$
$$p>4, \quad 2\tfrac{5}{11} < \beta_2 < 4\tfrac{1}{2};$$
$$p>6, \quad 2\tfrac{3}{5} < \beta_2 < 3\tfrac{3}{4}.$$

It will be found that for $p = 4$, the simpler form

$$I_x(p,p) \doteq \tfrac{1}{2} + \Phi(\zeta)$$

leads to a maximum error of 0·00045. For $p = 6$ this maximum error is halved.

2. Derivation of the approximation

The probability-density function of the median of a sample of $(2p-1)$ members from a population with distribution function $F(x)$ is proportional to

$$\{F(x)[1-F(x)]\}^{p-1} dF(x).$$

For a normal parent population Hojo (1931) has shown that this is very nearly normal for quite small p; for $p = 2$ he finds $\beta_2 = 3{\cdot}035$. Thus we have approximately

$$\{F(x)[1-F(x)]\}^{p-1} dF(x) \propto e^{-x^2/2\sigma^2} dx.$$

Thus the substitution $$t = F(\sigma\zeta)$$

will give the following approximation:

$$I_x(p,p) = \int_0^x t^{p-1}(1-t)^{p-1}\,dt \doteq \int_{-\infty}^{\frac{1}{\sigma}F^{-1}(x)} \frac{e^{-\frac{1}{2}\zeta^2}}{\sqrt{(2\pi)}}\,d\zeta.$$

It turns out to be better to use $$t = \tfrac{1}{2} + \left(\frac{\pi}{3}\right)^{\frac{1}{2}} \Phi\left(\left(\frac{3}{4p-1}\right)^{\frac{1}{2}} \zeta\right).$$

Expanding the function Φ about zero, we find that the integrand is proportional to

$$\left\{1 - \frac{(p-1)}{15(4p-1)^3}\zeta^6 - \frac{(p-1)}{35(4p-1)^4}\zeta^8 + \ldots\right\} e^{-\frac{1}{2}\zeta^2}.$$

Thus we have the approximation

$$I_x(p,p) \doteq \tfrac{1}{2} + K\left\{\Phi(\zeta) - \frac{(p-1)}{(4p-1)^3}\int_0^\zeta \frac{t^6}{15}\frac{e^{-\frac{1}{2}t^2}}{\sqrt{(2\pi)}}\,dt - \frac{(3p-1)}{(4p-1)^4}\int_0^\zeta \frac{t^8}{105}\frac{e^{-\frac{1}{2}t^2}}{\sqrt{(2\pi)}}\,dt\right\}.$$

Here K is a normalizing factor depending on p. Table 1 illustrates the behaviour of the incomplete sixth and eighth moments. These values do not differ greatly, and as the coefficients of these incomplete moments will be small, especially that of the eighth, we replace the latter by the sixth moment and derive

$$I_x(p,p) \doteq \tfrac{1}{2} + K\left\{\Phi(\zeta) - \frac{2(p-1)(2p+1)}{(4p-1)^4}\int_0^\zeta \frac{t^6}{15}\frac{e^{-\frac{1}{2}t^2}}{\sqrt{(2\pi)}}\,dt\right\}.$$

Table 1. *Values of standardized incomplete moments*

ζ	1·0	1·5	2·0	2·5	3·0	3·5	4·0
6th moment	0·00	0·03	0·11	0·24	0·37	0·45	0·49
8th moment	0·00	0·01	0·04	0·14	0·28	0·40	0·47

Since K will be very close to unity, this can be reduced to

$$I_x(p,p) \simeq \tfrac{1}{2} + L(p)\left\{\Phi(\zeta) - \frac{2(p-1)(2p+1)}{(4p-1)^4}\Psi(\zeta)\right\}.$$

We now choose $L(p)$ so that $I_1(p,p)$ takes the value unity.

3. Numerical data

Table 2, giving the values of $\Psi(\zeta)$, will be found to give sufficient accuracy, as two places in this function are quite adequate. More accurate values can be found from the tables of m_6 in *Tables for Statisticians and Biometricians*, vol. 1 (Pearson, K. 1914).

Table 2. *Values of* $\Psi(\zeta)$

ζ	·0	·1	·2	·3	·4	·5	·6	·7	·8	·9
0	0·000	0·040	0·079	0·118	0·156	0·191	0·226	0·258	0·288	0·315
1	0·339	0·360	0·377	0·391	0·400	0·406	0·406	0·403	0·395	0·383
2	0·367	0·348	0·326	0·302	0·272	0·249	0·222	0·196	0·171	0·147
3	0·125	0·105	0·087	0·071	0·057	0·046	0·036	0·028	0·022	0·017

The values of $L(p)$ can be found from Table 3.

Table 3. *Values of* $L(p)$

p	$L(p)$	p	$L(p)$
2·0	1·00090	3·0	1·00003
2·2	1·00054	3·2	1·00002
2·4	1·00023	3·4	1·00001
2·6	1·00012	3·6	1·00000
2·8	1·00006	3·8	1·00000

In order to check the accuracy of the formula, the function was evaluated to five places at some percentage points. The results are summarized in Table 4. The last line gives the maximum error of the simpler approximation,

$$I_x(p,p) \simeq \tfrac{1}{2} + \Phi(\zeta).$$

Table 4. *Errors of the approximation*

True value	$p=2$	$p=3$	$p=4$
1·0	0	0	0
0·995	+0·00002	−0·00005	−0·00002
0·99	+0·00002	−0·00006	−0·00003
0·975	+0·00003	−0·00007	−0·00003
0·95	+0·00006	−0·00007	−0·00003
0·90	+0·00011	−0·00006	−0·00001
0·75	+0·00004	−0·00003	−0·00001
0·50	0	0	0
Maximum error of simple form	−0·00205	−0·00085	−0·00045

The method used has been extended to $I(p, q)$ and has been found to give good results for p and q moderately close together. However, the loss of symmetry greatly complicates the work, and the approximation developed by Wise (1950) is much more useful.

4. Applications

The result for the Type II curve is

$$y = y_0\left(1-\frac{x^2}{a^2}\right)^m, \quad F(x) = I_X(m+1, m+1),$$

where $X = \frac{1}{2}(1+x/a)$.

For the Type VII curve we have

$$y = y_0\left(1+\frac{x^2}{a^2}\right)^{-m}, \quad F(x) = I_X(m-\tfrac{1}{2}, m-\tfrac{1}{2}),$$

where $X = \frac{1}{2}[1+x/(a^2+x^2)^{\frac{1}{2}}]$.

We note too that $I_x(p, \frac{1}{2})$ can be evaluated, since

$$I_x(p, \tfrac{1}{2}) = 2I_z(p, p), \quad 2z = 1-(1-x)^{\frac{1}{2}}.$$

For the distribution of 'Student's' t with n degrees of freedom,

$$\Pr(t<t_0) = I_X(\tfrac{1}{2}n, \tfrac{1}{2}n),$$

where $X = \frac{1}{2}[1+t_0/(n+t_0^2)^{\frac{1}{2}}]$.

A simple consequence of this expression does not seem to have been noticed. Let V denote the ratio of two independent estimates of variance, each based on n degrees of freedom. Then the statistic

$$\tfrac{1}{2}n^{\frac{1}{2}}(V^{\frac{1}{2}}-V^{-\frac{1}{2}})$$

has the t distribution on n degrees of freedom, the probability of the observed discrepancy being exceeded in random sampling corresponding to twice the tail area of the t curve.

The author wishes to thank Dr H. O. Hartley for some suggestions as to the presentation of this work. Acknowledgement is made to the Chief Scientist, Ministry of Supply, for permission to publish this note.

REFERENCES

Hojo, T. (1931). *Biometrika*, **23**, 315.

Pearson, K. (Ed.) (1914). *Tables for Statisticians and Biometricians*, **1**. University College, London: Biometrika Office.

Wise, M. E. (1950). *Biometrika*, **37**, 208.

The distribution of quantiles of small samples

By J. H. CADWELL

1. Introduction

It is well known that for any continuous population, the quantile of a sample of size n has an asymtotically normal distribution. Numerical investigations into the case of a normal parent population reveal the fact that, while the asymptotic mean and variance are only approached slowly, the convergence to normality is suprisingly fast.

E. S. Pearson & Adyanthaya (1928) evaluated the variance of the median for $n = 3, 4$ and 5. Hojo (1931) extended this up to $n = 12$, and evaluated β_2 for n up to 8. He also gave a number of results for quartiles. K. Pearson (1931*b*) developed another method which he applied to some special cases. Hojo fitted polynomials in $1/n$ to his values of the quantity:

$$c_n = \frac{\text{S.D. median of } n \text{ values}}{\text{S.D. mean of } n \text{ values}},$$

and obtained

$$n \text{ odd}: \quad c_n = \sqrt{\frac{\pi}{2}} - \frac{0{\cdot}2653}{n} - \frac{0{\cdot}0699}{n^2} + \frac{0{\cdot}0522}{n^3};$$

$$n \text{ even}: \quad c_n = \sqrt{\frac{\pi}{2}} - \frac{0{\cdot}8261}{n} + \frac{0{\cdot}7862}{n^2} - \frac{0{\cdot}3478}{n^3} + \frac{0{\cdot}1304}{n^4}.$$

In an appendix to Hojo's paper K. Pearson (1931*a*) derived the result

$$n \text{ odd:} \quad c_n = \sqrt{\frac{\pi}{2}} - \frac{0{\cdot}2690}{n} - \frac{0{\cdot}0782}{n^2} + \ldots.$$

This gives good agreement for $n = 11$, but there is a discrepancy at $n = 3$. Below we obtain approximate expressions for c_n and β_2 for the median. These are very close to Hojo's values down to the lowest value of n. The method used is also applied to the general quantile of a sample from a normal population.

While a normal population has been used throughout, the method could be applied to any continuous parent distribution without difficulty.

2. The median when $n = 2\nu + 1$

Since the median value exceeds ν values and is itself exceeded by ν values, its probability density will be

$$\frac{(2\nu+1)!}{\nu!\,\nu!}\{F(x)\,[1-F(x)]\}^{\nu}\,dF(x),$$

where $F(x)$ is the population distribution function.

By considering the integral $\iint \exp\{-\frac{1}{2}(x^2+y^2)\}\,dx\,dy$,

taken over the square with vertices $(\pm a, \pm a)$, and over a circle of equal area, we derive

$$\Phi(a) \doteq \tfrac{1}{2}\left\{1 - \exp -\frac{2a^2}{\pi}\right\}^{\frac{1}{2}}.$$

Here $\Phi(a)$ is the normal area from 0 to a. Using this result, the quantity

$$\exp -\frac{2\nu x^2}{\pi} \exp -\frac{x^2}{2}\,dx$$

is found to be approximately proportional to the probability density of the median.

While the expression used for $\Phi(a)$ is close for all values of a, I have shown (1951) that a modification of the equal areas condition leads to an improved result.

Adopting an alternative approach, and starting from the series expansion of $\Phi(x)$ about zero, the steps given below follow in order:

$$\begin{aligned} 4F(x)\{1-F(x)\} &= 1 - \frac{2}{\pi}\left\{x^2 - \frac{x^4}{3} + \frac{7x^6}{90} + \ldots\right\} \\ &= e^{-2x^2/\pi} + \frac{2(\pi-3)}{3\pi^2}\,x^4 + \frac{(60-7\pi^2)}{45\pi^3}\,x^6 + \ldots \\ &= e^{-2x^2/\pi}\left\{1 + \frac{2(\pi-3)}{3\pi^2}\,x^4 - \frac{120+7\pi^2-60\pi}{45\pi^3}\,x^6 + \ldots\right\}. \end{aligned}$$

It will be found that the coefficient of x^6 is $-0{\cdot}0004$. If this and higher powers be ignored, the median has a probability density proportional to

$$e^{-2\nu x^2/\pi}\left\{1 + \frac{2(\pi-3)\,\nu}{3\pi^2}\,x^4\right\} e^{-\frac{1}{2}x^2}\,dx.$$

We write this probability density in the form

$$\lambda\, e^{-\frac{1}{2}y^2}(1 + ky^4)\,dy,$$

where we choose λ so as to normalize the density, and put

$$y = \left(\frac{\pi+4\nu}{\pi}\right)^{\frac{1}{2}} x, \quad k = \frac{2(\pi-3)\,\nu}{3(\pi+4\nu)^2}.$$

After a little algebra we deduce

$$\mu_2(y) = 1 + 12k + O(k^2),$$
$$\beta_2(y) = 3 + 24k + O(k^2).$$

Hence we derive the approximate formulae

$$c_n \doteq \left\{\frac{\pi(2\nu+1)}{(\pi+4\nu)}\right\}^{\frac{1}{2}}\left\{1 + \frac{4(\pi-3)\,\nu}{(\pi+4\nu)^2}\right\},$$

$$\beta_2 \doteq 3 + \frac{16(\pi-3)\,\nu}{(\pi+4\nu)^3}.$$

Table 1 shows the degree of agreement with Hojo's computed values. The first term of the approximation to c_n, corresponding to the first approximation to $\Phi(a)$ given, gives for c_3 the value 1·149. If we expand the expression for c_n we obtain

$$c_n = \sqrt{\frac{\pi}{2}}\left(1-\frac{4-\pi}{2n}\right)+O\left(\frac{1}{n^2}\right),$$

which agrees with Pearson's formula quoted above. We note that Hojo's regression result is also quite close.

Table 1. *Median in odd samples*

n		c_n	β_2
3	Correct Approx.	1·1602 1·1615	3·0347 3·0444
7	Correct Approx.	1·2137 1·2141	3·0272 3·0296
11	Correct Approx.	1·2286 1·2286	— —

3. The median when $n = 2\nu+2$

In this case the median is defined conventionally as the mean of the $(\nu+1)$th value x_1 and the $(\nu+2)$th value x_2. The joint probability density for the variables x_1, x_2 is

$$\frac{(2\nu+2)!}{\nu!\,\nu!}F^{\nu}(x_1)\{1-F(x_2)\}^{\nu}dF(x_1)\,dF(x_2).$$

After a little manipulation we find that the probability density of the median is proportional to

$$e^{-x^2}dx\int_0^{\infty}F^{\nu}(x-t)\{1-F(x+t)\}^{\nu}e^{-t^2}dt.$$

This proves more troublesome to handle and we omit details. The results obtained are

$$c_n \simeq \left\{\frac{\pi(\nu+1)}{(\pi+2\nu)}\right\}^{\frac{1}{2}}\left\{1-\frac{(4-\pi)}{4(\pi+2\nu)}+\frac{(\pi-3)\,\nu}{(\pi+2\nu)^2}\right\},$$

$$\beta_2 \simeq 3+\frac{4(\pi-3)\,\nu}{(\pi+2\nu)^2}.$$

In deriving these results the substitution $y = \nu t$ was employed so that we do not obtain the correct value of c_n by putting $\nu = 0$.

Table 2. *Median in even samples*

n		c_n	β_2
4	Correct Approx.	1·0922 1·0647	3·0191 3·0216
8	Correct Approx.	1·1600 1·1509	3·0200 3·0204
12	Correct Approx.	1·1898 1·1844	— —

Table 2 compares the approximate values with those found by Hojo. Expanding c_n in powers of $1/n$ we get

$$c_n = \sqrt{\frac{\pi}{2}}\left\{1-\frac{6-\pi}{4n}\right\}+\ldots = \sqrt{\frac{\pi}{2}}-\frac{0{\cdot}8941}{n}+O\left(\frac{1}{n^2}\right).$$

Hojo's empirical result agrees quite well with this value. It will be seen that convergence of c_n is much slower in this case. By using another term inside the bracket it is improved, so that $c_{12} = 1{\cdot}1906$. The necessity for this term corresponds to the relatively large coefficient of $1/n^2$ in Hojo's relationship.

4. The general quantile

Consider a sample of $np+nq+1$ values with np integral and $p+q=1$. Then the probability density of the $(np+1)$th member of this sample will be

$$(n+1)\binom{n}{np}F^{np}(x)\{1-F(x)\}^{nq}\,dF(x).$$

Let the probability density of the parent population be expressed in a Taylor series about η, where $F(\eta)=p$, as

$$dF = c_0+c_1\zeta+c_2\zeta^2+\ldots.$$

Taking logarithms of the probability density of the quantile, we obtain

$$c+np\log\left(1+\frac{c_0\zeta}{p}+\ldots\right)+nq\log\left(1-\frac{c_0\zeta}{q}+\ldots\right)+\log\left(1+\frac{c_1\zeta}{c_0}+\ldots\right).$$

Expanding and collecting terms, this is found to reduce to

$$\frac{c_1\zeta}{c_0}-\left(\frac{nc_0^2}{2pq}+\frac{c_1^2}{2c_0^2}-\frac{c_2}{c_0}\right)\zeta^2+O(\zeta^3).$$

As n increases the sample quantile converges to the population value and ζ will be small. Hence ignoring the terms in ζ^3, etc., we find that the quantile is approximately normally distributed, with

$$\text{Mean:}\quad u=\eta\frac{nc_0^2}{(nc_0^2+pq)},$$

$$\text{Variance:}\quad v=\frac{pq}{(nc_0^2+pq)}.$$

Here we have replaced c_1 and c_2 by their values for a normal curve. We compare these values with the asymptotic formulae

$$\text{Mean}\quad =\eta,$$

$$\text{Variance}=\frac{pq}{(n+1)c_0^2}.$$

Table 3

n	7	11
p	$\frac{1}{8}$	$\frac{1}{4}$
Computed Values — Mean	0·757	0·729
Computed Values — S.D.	0·507	0·407
Approx. Values — Mean	0·705 (0·754)	0·700 (0·731)
Approx. Values — S.D.	0·518	0·412
Aspt. Values — Mean	0·966	0·842
Aspt. Values — S.D.	0·562	0·432

Table 3 shows the agreement with two cases taken from Hojo's paper. It will be seen that the approximate formulae provide results which are a big improvement on the asymptotic values. While further terms could be used, they become quite involved. The next approximation to the mean is found to be

$$u+\frac{nv^2c_0^2}{2p^2q^2}\{3pq\eta-2c_0(p-q)\}.$$

The values computed from this form are given in brackets in the table above. It is worth noting that in the first case Hojo computed

$$\beta_1=0{\cdot}026, \quad \beta_2=3{\cdot}088.$$

Thus the distribution of the sixth member in a sample of seven is surprisingly close to normality, although, as appears from the above table, the limiting values of mean and variance are very poor approximations to the true values.

The author wishes to thank Dr N. L. Johnson for some suggestions as to the presentation of this work.

Acknowledgement is made to the Chief Scientist, Ministry of Supply, for permission to publish this note.

REFERENCES

Cadwell, J. H. (1951). *Biometrika*, **38**, 475.
Hojo, T. (1931). *Biometrika*, **23**, 315.
Pearson, E. S. & Adyanthaya, N. K. (1928). *Biometrika*, **20A**, 356.
Pearson, K. (1931*a*). *Biometrika*, **23**, 361.
Pearson, K. (1931*b*). *Biometrika*, **23**, 364.

On a correction term in the method of paired comparisons

By J. A. van der Heiden, *Koninklijke/Shell-Laboratorium, Amsterdam*

The usual calculation according to the method of paired comparisons requires a correction term if observers decline to express a preference between a pair of objects. I shall show that this correction term is equal to one-quarter of the number of times of no-preference. In the method of paired comparisons the $\frac{1}{2}n(n-1)$ pairs out of a population of n objects are presented to m observers, who are supposed to express a preference for one member of each pair. If γ_{ab} is the number of observers having a preference for item a over item b,

and
$$\left.\begin{aligned}\gamma_{ab}+\gamma_{ba}&=m \quad (a\neq b),\\ \gamma_{ab}&=0 \quad\quad (a=b).\end{aligned}\right\} \tag{1}$$

If x observers have a preference for a over b and i observers decline to express preferences for a or b, γ_{ab} is defined by

$$\gamma_{ab}=x+\tfrac{1}{2}i. \tag{2}$$

Kendall & Babington Smith (1939) define a coefficient of agreement

$$u=\frac{2\Sigma}{\binom{m}{2}\binom{u}{2}}-1, \tag{3}$$

where
$$\Sigma=\sum_{a=1}^{n}\sum_{b=1}^{n}\binom{\gamma_{ab}}{2}, \tag{4}$$

which is a measure of agreement in appreciation in case the observers have always expressed a preference, i.e. when $i=0$.

If i is not zero, the terms $\binom{\gamma_{ab}}{2}+\binom{\gamma_{ba}}{2}$ in (4) have to be replaced by

$$\sum_{j=0}^{i}\binom{i}{j}2^{-i}\left\{\binom{x+j}{2}+\binom{m-x-j}{2}\right\}. \tag{5}$$

as the probability that γ_{ab} is $x+j$ $(0 \leqslant j \leqslant i)$ and $\gamma_{ba} = m-x-j$ is $\binom{i}{j} 2^{-i}$. It will be shown that the correction is equal to $\frac{1}{4}i$.

This correction is given by

$$c(x,m,i) = \sum_{j=0}^{i} \binom{i}{j} 2^{-i} \left\{ \binom{x+j}{2} + \binom{m-x-j}{2} \right\} - \left\{ \binom{x+\frac{1}{2}i}{2} + \binom{m-x-\frac{1}{2}i}{2} \right\}. \tag{6}$$

We first calculate

$$\sum_{j=0}^{i} \binom{i}{j} 2^{-i} \binom{x+j}{2} - \binom{x+\frac{1}{2}i}{2} = \frac{i}{4} - \frac{i^2}{8} - \frac{xi}{2} + 2^{-i}x \sum_{j=0}^{i} \binom{i}{j} j + 2^{-(i+1)} \sum_{j=0}^{i} \binom{i}{j} j(j-1). \tag{7}$$

Now
$$\sum_{j=0}^{i} \binom{i}{j} j = i \sum_{j=0}^{i} \binom{i-1}{j-1} = i 2^{i-1}$$

and
$$\sum_{j=0}^{i} \binom{i}{j} j(j-1) = i(i-1) \sum_{j=0}^{i} \binom{i-1}{j-1} = i(i-1)\, 2^{i-2}.$$

Substitution in (7) then gives
$$\sum_{j=0}^{i} \binom{i}{j} 2^{-i} \binom{x+j}{2} - \binom{x+\frac{1}{2}i}{2} = \frac{i}{8}, \tag{8}$$

and since x does not occur in this result we likewise find

$$\sum_{j=0}^{i} \binom{i}{j} 2^{-i} \binom{m-x-j}{2} - \binom{m-x-\frac{1}{2}i}{2} = \frac{i}{8}. \tag{9}$$

Hence
$$c(x,m,i) = \tfrac{1}{4}i. \tag{10}$$

This result is in agreement with the numerical results of Kendall (1948), calculated for two examples with $i = 1$ and $i = 2$.

REFERENCES

KENDALL, M. G. (1948). *Rank Correlation Methods.* London: Charles Griffin and Co.
KENDALL, M. G. & BABINGTON SMITH, B. (1939). *Biometrika*, **31**, 324.

REVIEWS

Chance and Choice by Cardpack and Chessboard. By L. HOGBEN. Max Parrish & Co. Ltd. Pp. i+417. 50*s*.

It is claimed that 'Choice and Chance' aims to make the logical basis of statistical theory, that is the theory of probability, accessible to any college student who has the intellectual equipment to make use of statistical methods or to benefit from a practical course, and the author claims that this is done against the background of the new theory of logical inference developed in America by Neyman's School (!). What Professor Hogben has done is to write a textbook of statistical methods in which the proofs of probability theorems are magnificently illustrated by coloured pictures.

Among the subjects covered are permutations and combinations, expectations and moments, distribution of a proportion, confidence intervals, and correlation. The approach to the subject is not particularly novel, and once the statistician has grasped Professor Hogben's nomenclature and symbolism he will find that there is little here which he has not read before. Statisticians appear, broadly speaking, to be divisible into two categories—those who find a diagram helpful and those who do not. The former class are appreciative of R. A. Fisher's geometrical approach to sampling distributions and are undoubtedly in the majority. This being so there will be many who will like to read this book, and possibly be stimulated by it.

While Prof. Hogben may stimulate the professional statistician by his book, it is to be doubted whether he will succeed in teaching the biologist and non-mathematical workers generally. Euclid is reported to have remarked that there is no royal road to learning and any non-mathematician who has patience to puzzle out Prof. Hogben's diagrammatic approach would probably find his time better employed with an old and tried classic such as Yule and Kendall's *Introduction to the Theory of Statistics*. The style of writing is irritating and at times unnecessarily florid. For example (p. 268), it is to be questioned whether the student's knowledge of moment generating functions is increased by such explanations as 'Having witnessed the process of parturition specified by (xv) which delivers the function (f_x) of x out of the womb of $G_t(f_x)$ etc.' Such remarks as (p. 256) that 'it will not be necessary to traverse the tortuous steps which led Pearson himself to discover the general pattern of his system' will only have sense if the author is substituting something less tortuous himself, and so on. However, Prof. Hogben is clearly an enthusiast, and as such much may be forgiven him.

The writer found a curious archaic echo in such remarks as 'Since it is the fashion to decry that part of statistical theory which rightly pertains to large samples etc.' Years ago we had large samples with K. Pearson and small samples with R. A. Fisher. The writer thought that nowadays statisticians just had samples, all of which (given the same initial assumptions) could be treated by the same exact techniques. The interpretation of our techniques we shall always wrangle over, but since the final interpretation of a probability is a subjective business, this wrangling is possibly inevitable.

The book is excellently printed and produced. Its price, however, is likely to put it beyond the reach of many of those at whom it is aimed.

F. N. DAVID

Hypothesis-testing in Time Series Analysis. By PETER WHITTLE. Uppsala: Almquist and Wiksells. 1951. Pp. 120. 18*s*.

In the present state of development of statistical theory it is taken as axiomatic by some statisticians that no test of significance of a statistical hypothesis should be constructed unless the alternatives to the hypothesis tested are clearly set out. There are exceptions to this statement, of course, the χ^2 test being the classic example. Nevertheless, in general, if a sensitive test is desired, the test is usually constructed so that a particular class of alternative hypotheses will not be overlooked. Simple illustrations of this are found in tests for randomness. We cannot define with any exactitude in a single comprehensive statement all the facets of randomness, but we can define what we mean by a particular non-random departure. Hence in testing for randomness we select from the numerous possible tests that particular one which will be most sensitive to that type of departure from randomness we are most anxious not to overlook. The battery of tests for randomness is useful statistically simply because the various alternative hypotheses have been stated.

If we accept the point of view that a statement of the hypotheses alternate to that tested is a prerequisite to the construction of a test then, in spite of the considerable amount of research which has been done on time series analysis, the application of the theory can only be described as being in a thoroughly unsatisfactory state. The type of analysis carried out appears to depend to a large extent on the whim of the person applying the statistical method, and while any test may be better than no test, the underlying suggestion is (almost) that any test will do. It is this nebulous situation which Mr Whittle has set himself to clarify. He endeavours, for example, to construct tests of the hypothesis that the data follow an autoregressive scheme with the alternative that if they do not, then they follow a moving average and vice versa. The solutions are only approximate and it is a little difficult to see what would be the effects if the several assumptions which are made are not satisfied. However, it is more than sufficient to say that Mr Whittle is a pioneer and it has always been the fate of pioneers both to stimulate those who follow and to be criticized by those who are wise after the event. All who are interested in time series will benefit by reading this book if only from the stimulation and excitement which comes from trying to go one better than the author. This is an important contribution to the research work on time series and may well prove to be the foundation stone of a satisfactory theory.

Having said this the reviewer may perhaps be permitted to grumble a little. It is pertinent to remark on the air of detachment from and seeming unawareness of a certain amount of statistical mathematics (e.g. §44 on the distribution of quadratic functions) which is well known enough to take for granted. It is surmised that this book is a doctoral thesis in which 'padding' to a greater or less degree is to be expected. It is unfortunate, however, for the result of such a procedure makes it possible for one to overlook the author's view and important ideas in the restatement of those of other people.

Because the book is so important the real criticism which the reviewer would make is that it is to be deplored that the new techniques stand or fall by means of specious arguments based on Bayes's theorem. The author starts off with Bayes's theorem and as is inevitable, finds himself in the position of having to make assumptions regarding the probability of hypotheses. Making arbitrary assumptions and proceeding to a limit he arrives at what is in fact the likelihood ratio, when he finds, not surprisingly, that the application of his disguised likelihood ratio to certain classical problems produces classical results. It would have seemed simpler to skip Bayes and to start off with the likelihood ratio as a principle. However, it is always easier to be critical than to create, and research workers for some time to come will be grateful to Mr Whittle for introducing them to an interesting set of new ideas.

F. N. DAVID

Foundations of the Theory of Probability. By A. N. KOLMOGOROV. Pp. v + 71. $2.25.

Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. By A. KHINTCHINE. Pp. 76. Chelsea Publishing Company, N.Y. $2.00.

Probability theory in the past fifty years has owed its development principally to the French and the Russian schools. Many probabilists are familiar with the work of the French, and others have been made aware of the extent of the Russian contribution by reading the work of Cramér, of Uspensky and of Feller. In the two books published by the Chelsea Company we have reprinted works on probability which must have been authoritative at the time at which they were written (1933).

Prof. Kolmogorov gives a summary of the axiomatic approach to the theory of probability. Nothing is contained here which has not been repeated and extended in the publications edited by M. Borel, but the development is clear, vigorous and satisfying and will be enjoyed by all who like the mathematics of probability.

Prof. Khintchine discusses stochastic processes, and his work gives a picture of the elementary development which will be found useful even at the present time, although there is little here which has not been covered by Feller in recent publications.

For the student of probability who finds fresh understanding of theory with a fresh point of view, the return to beginnings as represented by these reprints will prove a fruitful pilgrimage. Those just starting seriously to study probability will find these books most helpful.

It is unfortunate that it should still be considered necessary to publish Prof. Khintchine's book 'by Authority of the Attorney General of the United States, etc.' It is difficult to see that this book could ever have been regarded as other than pure science. Further, an author should receive royalties for his work.

F. N. DAVID

Statistical Methodology Reviews, 1941–1950. Edited by OSCAR KRISEN BUROS. x+457 pp. New York: John Wiley and Sons. $7.00; London: Chapman and Hall. 56*s*.

This handsomely produced collection of reviews includes extracts from fifteen reviews of its predecessor, which reveal the general appreciation of the earlier work of Prof. Buros. This later volume, confined more strictly to methodological statistical books, can hardly be less appreciated by the statistical community. Its main value will be as a preliminary guide to anyone who has to choose between text-books he does not know; an inspection of these reviews will sometimes shorten the list of books which must actually be examined. By professional statisticians, and in particular the post-war generation, the book will also be used to satisfy justifiable curiosity concerning reviewers. Despite the greater selectivity of his latest work, Prof. Buros has still provided us with ample opportunity to compare individual judgments, and, occasionally, to indulge in a little *Schadenfreude*.

A. STUART

Editorial Note.

It has been brought to the Editors' attention that the result obtained in the paper 'Note on the inversion theorem' by J. Gil-Pelaez (*Biometrika*, December 1951, pp. 481–2) is a special case of results obtained in the paper 'Inversion formulae for the distribution of ratios' by John Gurland (*Annals of Mathematical Statistics*, June 1948, pp. 228–37).

(*All Rights reserved*)

BIOMETRIKA. Vol. 39, Parts 1 and 2

CONTENTS

	PAGE
Moment coefficients of the k-statistics in samples from a finite population. By JOHN WISHART	1
Moment-statistics in samples from a finite population. By M. G. KENDALL	14
Some exact tests in multivariate analysis. By E. J. WILLIAMS	17
The construction of balanced designs for experiments involving sequences of treatments. By H. D. PATTERSON	32
Multi-factor designs of first order. By G. E. P. BOX	49
Tests of significance in canonical analysis. By F. H. C. MARRIOTT	58
The interpretation of interactions in factorial experiments. By E. J. WILLIAMS	65
On sampling from a population of rankers. By A. S. C. EHRENBERG	82
Least-squares estimation of location and scale parameters using order statistics. By E. H. LLOYD	88
Regression, structure and functional relationship. Part II. By M. G. KENDALL	96
On the concurrence of a set of regression lines. By K. D. TOCHER	109
A sampling test of the χ^2 theory for probability chains. By M. S. BARTLETT	118
On mathematical analysis of style. By WILHELM FUCKS	122
Comparison of two approximations to the distribution of the range in small samples from normal populations. By E. S. PEARSON	130
The covering circle of a sample from a circular normal distribution. By H. E. DANIELS	137
The frequency justification of certain sequential tests. By G. A. BARNARD	144
Experimental designs for serially correlated observations. By R. M. WILLIAMS	151
The time intervals between industrial accidents. By B. A. MAGUIRE, E. S. PEARSON and A. H. A. WYNN	168
The estimation of death-rates from capture-mark-recapture sampling. By P. A. P. MORAN	181
MISCELLANEA	
A note on the design problem. By K. D. TOCHER	189
Tables of percentage points of the 'Studentized' extreme deviate from the ample mean. By K. R. NAIR	189
Extended and corrected tables of the upper percentage points of the 'Studentized' range. by JOYCE M. MAY	192
On the distribution of 'Studentized' range. By K. C. S. PILLAI	194
Note on a certain family of discrete distributions. By J. S. MARITZ	196
Some properties of runs in smoothed random series. By ALISON M. GRANT	198
An approximation to the symmetrical incomplete beta function. By J. H. CADWELL	204
The distribution of quantiles of small samples. By J. H. CADWELL	207
On a correction term in the method of paired comparisons. By J. A. VAN DER HEIDEN	211
REVIEWS	
L. HOGBEN's 'Chance and Choice by Cardpack and Chessboard'	213
P. WHITTLE's 'Hypothesis-testing in Time Series Analysis	213
A. N. KOLMOGOROV's 'Foundations of the Theory of Probability'	214
A. KHINTCHINE's 'Asymptotische Gesetze der Wahrscheinlichkeitsrechnung'	214
O. K. BUROS's (ed.) 'Statistical Methodology Reviews'	215
EDITORIAL NOTE	

First printed in Great Britain at the University Press, Cambridge
Reprinted by offset-litho by Percy Lund, Humphries & Co., Ltd, Bradford